Quantifying functional imaging modalities and their application in the assessment of treatment response and prognosis in bone metastases

Taylor, Benjamin Peter

Awarding institution:
King's College London

Quantifying functional imaging modalities and their application in the assessment of treatment response and prognosis in bone metastases

Dr Benjamin Peter Taylor

MBChB BMSc (Hons) MRCP FRCR PGDip (Oncology)

June 2017

Department of Cancer Imaging
Division of Imaging Sciences & Biomedical Engineering

A thesis submitted to

King's College London

for the degree of

Doctor of Medicine (Research)

Abstract

Bone metastases are common in malignancy and cause morbidity. However, assessing treatment response of bone metastases with morphological changes is insensitive, variable and lacks specificity. This thesis reports initial recruitment data from three prospective trials, examining the potential role of functional imaging methods as early biomarkers of response and prognosis. Breast and prostate cancer patients with bone metastases had ^{18}F-Fluoride Positron Emission Tomography (PET) and Diffusion-Weighted Magnetic Resonance Imaging (DW-MRI) before treatment and again at 8-12 weeks following treatment commencement. Clinical response outcomes were appraised up to 24 weeks. For both imaging methods, there were significant differences in the parameters between breast and prostate metastases reflecting different bone metastasis pathophysiology of these tumours, and significant intra-patient inter-lesion heterogeneity was evident. ^{18}F-Fluoride per-lesion analysis (62 lesions) identified: an increase in SUV_{max} ($p=0.048$predicted for treatment response, a lower $SUV_{mean/max/peak/entropy/Ki}$ and higher SUV_{energy} before treatment were significantly associated with OS benefit (all $p<0.05$) and higher $\%\Delta SUV_{mean/max/peak/entropy/Ki,}$ and lower $\%\Delta SUV_{energy}$ at 8 weeks were significantly associated with an OS benefit (all $p<0.05$). DW-MRI per-lesion analysis (92 lesions) identified a higher %ΔADC predicted for treatment response ($p=0.029$) and a lower baseline ADC is prognostic of OS ($p<0.05$). Per-patient analyses (^{18}F-Fluoride PET n=12; DW-MRI n=20) yielded less predictive and prognostic value, and have demonstrated the significant impact of VOI definition methods on the ^{18}F-Fluoride quantification. Whole-body methods have been developed for DW-MRI and ^{18}F-Fluoride PET to enable quantification of the heterogeneous skeletal metastatic burden; whole-body parameters were not more predictive or prognostic than per-patient analyses. DW-MRI and ^{18}F-Fluoride PET have shown potential as early imaging biomarkers of treatment response and prognosis for bone metastases in breast and prostate cancer, and significant methodological factors have been considered, aiming to achieve practical clinical utility of these techniques.

Table of Contents

Table of Figures

Table of Tables

Acknowledgements

I would like to express my sincere gratitude to my MD supervisor, Professor Gary Cook, for his unrivalled expertise and limitless patience. This project would not have been realised without his wisdom, experience and constant support. I would also like to thank by second supervisor, Dr Simon Hughes, for his pragmatic advice, insight and encouragement, and for his significant help as a colleague in recruiting patients. Professor Vicky Goh has been very generous with her time and knowledge, and I am very grateful for her guidance on MR imaging. I was fortunate to work alongside Dr Musib Siddique who developed the software so essential to this thesis, and whose help was considerable. I would also like to thank Dr Connie Yip, my companion research fellow, who provided great friendship and support.

All of the staff of the PET Centre at St Thomas' Hospital, London, made me feel part of the team, and it has been a pleasure working alongside them. I am indebted to all my colleagues at Guy's Cancer Centre too numerous to individually name who have guided me at the start of my consultant career and inspired me – and kept me sane.

I am to grateful to Alliance Medical, and in particular Dr Charles Niehaus, for the funding that made my research fellowship possible.

I have, particularly recently, neglected my family, extended family, friends, and especially Simon. I want to thank them all for their love and understanding.

Abbreviations

Abbreviation	Meaning
%Δ	Percentage change
^{18}F-FDG	^{18}F-labelled fluorodeoxyglucose
3D	Three dimension
95%CI	95% Confidence interval
^{99m}Tc-MDP	^{99m}Tc labelled Methylene diphosphonate
ADC	Apparent Diffusion Coefficient
ANOVA	Analysis of Variance
ARSAC	Administration of Radioactive Substances Advisory Committee
AUC	Area under the curve
B/BCa	Breast cancer
Bq	Becquerel
CMR	Complete metabolic response
CR	Complete response
CT	Computed Tomography
CTX	C-telopeptides
d.f.	Degrees of freedom
DCI-MRI	Dynamic Contrast Enhanced MRI
DWI	Diffusion weighted imaging
DW-MRI	Diffusion Weighted Magnetic Resonance Imaging
EORTC	European Organisation for Research and Treatment of Cancer
F	F statistic
FAB-B	Functional Assessment of Bone metastases – Breast
FAB-IE	Functional Assessment of Bone metastases – Integrin Expression
FAB-P	Functional Assessment of Bone metastases – Prostate
Fig.	Figure
Figs.	Figures
FLAB	Fuzzy local adaptive Bayesian
FOV	Field of view
G-CSF	Granulocyte-colony Stimulating Factor
HR	Hazard ratio
IDF	Image derived function
IF	Input function
keV	Kiloelectronvolt
Ki	Input function
KM	Kaplan Meier
MDA	MD Anderson
MR	Magnetic Resonance
MR	Metabolic response
MRI	Magnetic Resonance Imaging
MTV	Metabolic Tumour Volume
n=	Number=
η^2	Eta squared statistic
NASA	National Aeronautics and Space Administration
NEX	Number of excitations
NRES	National Research Ethics Service
NSCLC	Non-small cell lung cancer
OS	Overall survival
P/PCa	Prostate cancer
PCWG2	Prostate Cancer Working Group 2
PD	Progressive Disease
PET	Positron Emission Tomography
PET/CT	Positron Emission Tomography/Computed Tomography
PFS	Progression free survival
PMD	Progressive metabolic disease
PMR	Partial metabolic response
PR	Partial response

PSA	Prostate specific antigen
PSA	Prostate specific antigen
RANK	Receptor activator of nuclear factor kappa-B
RANK-L	Ligand to RANK
RCP	Ethics committee
REC	Research Ethics Committee
ROC	Receiver Operator Curve
ROI	Region of interest
ROI^{CT}	ROI defined on
ROI^{PET}	ROI defined on PET
SMR	Stable metabolic response
SNR	Signal-to-noise ratio
SPECT	Single Photon Emission Computed Tomography
ST(mm)	Slice thickness(mm)
SUL	SUV normalised to lead body mass
SUV	Standardised uptake value
T	Tesla
t	t-statistic
Tbl.	Table
Tbls.	Tables
tDV	Total diffusion volume
TE(ms)	Time to echo in milliseconds
TGF-β	Transforming growth factor beta
TL	Target Lesion
TLA	Total Lesion Activity
TR(ms)	Repetition time
UICC	International Union Against Cancer
USA	United States of America
VOI	Volume of interest
WB	Whole-body
WHO	World Health Organisation
χ^2	Chi squared statistic

Chapter 1 Introduction

1.1 Bone Metastases

Bone is a common site for metastatic spread, and results in significant morbidity. Advances in therapies means patients survive longer with such disease. Breast and prostate cancers account for 80% of metastatic bone disease[1], both causing bone metastases in 65-75% of patients[1]. 85% of breast cancer patients at autopsy have bone metastases[2], exceeding the incidence identified during life with conventional imaging.

Different cancers have different predilection for bone involvement highlighting the significance of the underlying pathophysiology, including molecular and cellular biological characteristics, both of the metastatic cells and the tissues they spread to[3]. This was recognised by Paget in his 1889, 'seed and soil' hypothesis[4]. Most metastatic cells spread to bone via the blood[5]. Axial skeletal bones contain most of the blood-rich red bone marrow were 90% of deposits are identified[6], providing a large blood supply and the opportunity for 'blood stagnation'[5].

Normal bone is in a dynamic state of resorption and remodelling, coordinated by osteoclasts and osteoblasts[7]. Osteoclasts produce proteolytic enzymes and hydrogen ions in the localised environment adjacent to their ruffled border[7]. Coupled to bone resorption is new bone formation by osteoblasts, which produce the proteins of the bone matrix (primarily type I collagen)[7] and alkaline phosphatase, which are clinical markers of osteoblastic activity. This osteoid matrix is mineralised with hydroxyapatite crystals.

Once metastatic cells have become established, factors leading to osteoclast-mediated bone destruction are secreted. Receptor activator of nuclear factor kappa-B (RANK) and its ligand (RANKL) are the final common mediators for osteoclast maturation and activation in metastases. Growth factors (including transforming growth factor beta (TGF-β), prostaglandin-E, epidermal growth factor, tumour necrosis factor and interleukin-1[7]) are then released from the damaged bone matrix, further stimulating tumour cell growth, and potentiating the secretion of more bone-resorbing factors, enhancing osteolysis – the 'vicious cycle' of bone metastases[8]. TGF-β enhances osteoclast formation from precursor cells, inhibits osteoblast differentiation (thereby inhibiting new bone formation), and promotes the expression of parathyroid hormone related peptide which further enhances bone invasion by the tumour[9]. Osteoclast involvement is observed in all types of metastatic bone deposit, whether lytic or blastic. Raised levels of RANKL causing cancer-induced osteoclastogenesis result from both the infiltrating immune cells and reactive changes within bone, but also RANKL expression has been reported in several tumour types, including breast and prostate cancer[10].

Metastatic bone deposits can be primarily osteolytic (50%[11]), sclerotic (35%[11], particularly in prostate cancer deposits) or mixed (15%, particularly breast cancers[5]). Osteoblast growth factors (including TGF-β and platelet-derived growth factor) have been isolated from prostate tumours[7], and goes some way to explain this particular pattern of bony metastatic spread. More anaplastic

tumours, rapidly growing or those confined to the bone marrow tend to cause less osteoblastic stimulation[5], and lytic lesions result.

1.1.1 Prostate Cancer Bone Metastases

Prostate cancer is the most common cancer affecting men in the developed world and the incidence is rising. In the UK there were 40,460 new cases of prostate cancer diagnosed each year between 2008 and 2010[12], forming nearly 26% of male cancers (excluding non-melanoma skin cancers) in England, similar to the incidence of 27% in the USA[13]. The age-specific mortality and absolute number of deaths from prostate cancer have both been decreasing since 1993[14].

Prostate cancer spreads to the bone in approximately 90% of patients with advanced prostate cancer[15]. The preference for human bone by prostate cancer metastases has been demonstrated in animal models[16]. Prostate cancer metastases preferentially attach to the endothelium of bone marrow through the interaction of chemo-attractants[17], including stromal-derived factor-1[18].

Koeneman et al.[19] raised a hypothesis of 'oestomimetic' properties of prostate cancer cells, suggesting prostate cancer cells exhibit properties similar to the native bone cells, resulting in dis-regulation of the normal bone turnover. Despite the sclerotic clinical findings, prostate cancer bone metastases result in increased activity of both osteoclasts and osteoblasts. Bisphosphonates have been successfully used to inhibit osteoclast bone degradation, and proven reduction in skeletal related events (including pathological fracture)[20] in prostate cancer patients. Serum markers of bone degradation (including type I collagen C-telopeptides (CTX)) which demonstrate enhanced osteoclastic activity are significantly higher in prostate cancer patients[21] Bisphosphonate therapy is effective in reducing bone pain in patients with hormone refractory prostate cancer[22], suggesting morbidity results from osteoclastic action rather than the more clinically evident osteoblastic driven sequelae.

1.1.2 Breast Cancer Bone Metastases

Breast cancer is the most common malignancy in women in the UK, with 41523 new diagnoses in 2011[23]. Up to 75% of patients with metastatic breast cancer have bone metastases[2,24,25], with isolated bone metastases seen in around 20% of patients[26]. Osteolytic lesions reflect increased bone resorption by the osteoclasts, but the osteoclast formation and activity is regulated by the osteoblast[27].

Several bone targeted treatments are now approved for breast cancer patients with bone metastases, and are considered as standard of care[28-30]. These treatments include bisphosphonates and RANKL inhibitors. Bisphosphonates induce apoptosis of osteoclasts, thereby inhibiting bone resorption[31]. The RANKL pathway is responsible for the stimulation of osteoclasts by osteoblasts; osteoclast precursors express the cell surface receptor RANK which can be bound by RANKL expressed on activated osteoblasts. This interaction is targeted by anti-RANKL therapies, including Denosumab, a monoclonal antibody targeting the receptor activator of RANK-L, preventing activation of RANK on the osteoclast surface. Denosumab therapy in breast and prostate cancer reduces bone resorption, tumour induced bone destruction, and skeletal related events[32,33].

1.1.3 Imaging Bone Metastases

Bone metastases disrupt the balance between osteoclastic and osteoblastic activity, resulting in sclerotic or lytic bone lesions; the effect on bone architecture is identified with conventional imaging methods.

1.1.3.1 Plain X-Rays

Plain X-rays are frequently used for the initial skeletal assessment. Plain X-rays are relatively insensitive (44-50% less sensitive than skeletal scintigraphy[34] for breast cancer bone metastases) and only have a limited role in monitoring asymptomatic metastases[35], particularly in trabecular bone disease with the limited contrast between benign and malignant tissue. 50-70% of bone must be lost to be visible on simple plain films[36,37] Typically, osteolytic bone lesions will result in thinning of the trabeculae with an ill-defined margin on plain X-ray imaging[38]. Sclerotic disease causes nodular, rounded and commonly well circumscribed lesions secondary to trabeculae thickening[38].

Sclerosis is a likely outcome of a treatment response. This process starts at the periphery of a lesion and progresses towards the centre of a lesion[38]. However, it may take many months for this change to be identified on plain films (typically 3-6 months).

Plain films still have an important role for identification and assessment of bone metastases, but largely to provide confirmatory information for other imaging modalities, or to assess fracture risk.

1.1.3.2 CT

Computed tomography (CT) detects bone metastases with a sensitivity of 74% and specificity of 56%. Such imaging has advantages over plain-film assessment, with increased contrast, and the ability to apply bone-specific algorithms and window widths and levels, the ability to view images in multiple planes, and utilisation of the contrast between bone cortex and marrow. CT can often demonstrate bone marrow metastases before cortex bone destruction is demonstrable[39].

1.1.3.3 MRI

Magnetic Resonance Imaging (MRI) is a term encompassing a range of imaging techniques utilising on the phenomenon of nuclear magnetic resonance. Nuclei of atoms, when exposed to a magnetic field, absorb and re-emit electromagnetic radiation. The energy absorbed and re-emitted by a nucleus is influenced by the atomic isotope, the local environment of the nucleus, and the magnetic field strength. Not all nuclei resonate at precisely the same frequency because the immediate magnetic environment is affected by surrounding subatomic particles, particularly electrons; electrons can shield the atomic nuclei from the external applied magnetic field. This discrepancy is exploited to differentiate between tissues.

For clinical magnetic resonance imaging, a static magnetic field within the scanner causes hydrogen nuclei (protons) within the subject to line up along the magnetic field (often referred to as the B_0 field). These protons will be aligned directly with the field, or directly counter-aligned (i.e. at 180º), but all along the z-axis (along the B_0 field). This process is not quite random, and

there is an unequal distribution between these two alignments. There is a slight bias towards protons being aligned in the direction of the B_0 field, resulting in a net magnetisation in the direction of the magnetic field. A radio frequency pulse (B_1 field) to match the natural resonant frequency of protons is applied perpendicular to the B_0 field. This energy is absorbed, raising the nuclei to a higher energy state, which knocks the protons out of alignment with the B_0 magnetic field, and the magnetisation vector rotates by 90^0 into the x-y axis. This shift in the magnetisation can be detected by the MRI scanner.

The protons, knocked out of alignment with the primary magnetic field, return to their original positions once the radio-frequency pulse has passed. This process is termed relaxation, and is a measure of how quickly the net magnetisation vector returns to its ground state. There are several ways the relaxation process can be described.

T1 relaxation time is a measure of how fast the net magnetism vector returns to the original state in the direction of the B_0 field. During this process the previously excited nuclei return to their low energy state, and the energy they gained is lost to the surrounding tissues.

When the radio-frequency pulse knocks the protons into rotation in the x-y axis, not all the protons spin at the same speed; they are out of phase. T2 relaxation describes how the protons relax back into phase along the x-y axis. This happens quicker than the T1 relaxation.

Different tissues exhibit different proton densities and have different T1 and T2 relaxation characteristics. Images can be taken emphasising these characteristics by altering variables of the scan, including the time between radio frequency pulses. The more time between pulses, the more T1 or longitudinal relaxation is achieved, emphasising T1 signal. T2 weighed images are gained by changing the echo time, which refers to the time between the radio frequency pulse and the MR signal sampling.

On T1-weighted images, bone metastases can be identified as hypo-intense lesions against a background of marrow which return a higher signal intensity. On a T2-weighted sequence, bone metastases usually demonstrate T2 hyper-intensity due to their elevated water content[38].

MRI is not without limitations. Cortical bone has a short T2-relaxation time and is poorly characterised with MRI. Bone with a low marrow content, such as ribs, might be imaged better with CT[39].

1.1.3.3.1 Diffusion-Weighted MRI

The resonance frequency of protons is dependent on the strength of the magnetic field. If a sample is placed in a magnetic-field gradient, the protons within the sample will display a range of resonance frequencies. This characteristic is exploited to enable measurement of the ability of water to move freely within a sample.

Free water molecules have kinetic energy, related to temperature. This energy, translated into movement through thermal convection currents, causes the Brownian motion of particles suspended in water. On the atomic level this means water molecules are always on the move. If there are barriers preventing free movement, water molecule will move less far over a given time

than if the molecule has no obstruction to free movement. At 37°C the average displacement of water molecules is 0.6micrometers/ms[40]. It is this freedom of random motion of water content in tissues measured through 'apparent diffusivity', or diffusion weighted imaging, by measuring the effective displacement of water molecules over a given time period[41]. This can be quantified using MR imaging techniques.

Whereas more simple MR imaging relies on one magnetic field (B_0) and a radio frequency pulse applied at 90°, diffusion-weighted imaging requires additional magnetic fields, applied as non-uniform gradients. The first of these alters the phase of each proton of water molecules, but the amount of phase shift is relative to each proton's position within the B_0 field. The second additional gradient pulse is equal in magnitude, but opposite in direction, to the first. If the proton has not moved between these two pulses the net effect will be a complete reversal of the original phase shift, termed a 're-phasing', and a full signal is received by the MRI scanner. However, if the proton has moved in-between the first and second pulse gradients then complete re-phasing is not achieved; the degree of re-phasing is proportional to the magnitude (and the direction) of movement of the proton.

Where water can move freely one might anticipate more signal loss than where water movement is restricted in some way, and this is dependent on the architecture of the tissue being examined[42]. Cancer deposits generally have a higher cell density then benign and normal tissues, and can have a higher water and protein content[43]. Cancer tissues tend to exhibit restriction of water diffusion. Other factors impacting water diffusivity within tissues include cell density, cell nuclear:cytoplasm ratio, integrity of the cellular membranes and the extra-cellular space architecture[43], the presence of necrosis, and the architecture of the extracellular extravascular space.

The image sensitivity to diffusion is commonly referred to as the *b*-value. This *b*-value can be altered by changing the time between the two additional gradient pulses, and by altering the size of the gradient. The greater the *b*-value, the more sensitive the imaging to diffusion, and smaller diffusion differences are detected. Signal-to-noise ratio (SNR) and the T2-relaxation rates limited the interval for diffusion measurement usually within 40-80ms[41]. An apparent diffusion coefficient (ADC) map is generated using different *b*-values, typically including one with no diffusion weighting. By combining these maps, it is possible to create an ADC map free of the influencing factors other than the movement of water.

DW-MRI is very sensitive to movement artefact. Small movements caused by pulsatile arteries/arterioles can interfere with measurements, and there are larger patient factors including breathing, peristalsis and muscular movements[42]. Other artifacts result from areas of large variations in magnetic susceptibility, including airspaces in the lungs. Fat and soft tissue interfaces can result with artifacts, but modern DW-MRI imaging is always performed with fat-suppression, which limits the impact from this. The skeleton, however, is likely to be less affected by organ movement, and by areas of large magnetic field inhomogeneities.

1.1.3.3.2 Detection of bone metastases with MRI

MRI offers an advantage because it is sensitive to early changes within the bone marrow, changes occurring before the osteoclastic/osteoblastic response of the bone matrix[44] imaged with the other techniques.

MRI techniques have diagnostic advantages over other imaging techniques for characterisation of skeletal metastases[45-52] and has utility for complementary information to standard scintigraphic skeletal evaluation, identifying different lesions to those detected with scintigraphy[53]. Technical advances in the use of diffusion-weighted MRI imaging have led to its increased use for characterising bone metastases. Calculation of ADC values allows differentiation of benign from malignant tissue, e.g. head and neck tumours[54] and breast tumours[55-57]. Byun et al.[58] recorded the ADC of spinal metastases, and generated a method to use serial ADC measurements to monitor radiotherapy response. They reported bone metastases with an ADC of $0.78 \pm 0.03 \times 10^{-3}$ mm^2/s compared with $0.33 \pm 0.03 \times 10^{-3}$ mm^2/s of normal marrow. Following radiotherapy treatment, the group reported the ADC increased to $1.22 \pm 0.02 \times 10^{-3}$ mm^2/s, when reviewing the 23 patients with a symptomatic benefit from radiotherapy; T1 and T2 MRI imaging of the same lesions did not identify any changes, suggesting the benefit of using ADC quantification to monitor therapy responses[58].

1.1.3.4 Bone Scintigraphy

^{99m}Technetium (^{99m}Tc)-based bone-specific bisphosphonate-related radiopharmaceuticals, bind to the hydroxyapatite crystals mineralising the osteoid matrix created by osteoblasts in response to bone damage and normal bone turnover. This binding is proportional to the local blood flow, the osteoblastic activity[59] and the efficiency of extraction[38]. Imaging changes in osteoblastic metabolism confers clear advantages over cross-sectional imaging; ^{99m}Tc-MDP bone scintigraphy is sufficiently sensitive to detect a 5% change in osteoblastic turnover[60], whereas a 50% loss of bone mineral may be required for detection on radiographic imaging[60]. This translates into a sensitivity of ^{99m}Tc-MDP scintigraphy of between 70% and 100%[60], depending on the type and site of the tumour. Talbot et al. report a sensitivity of 78% and a specificity of 48%[61].

Planar bone scintigraphy can result in false-negative scans of 50%[62], particularly for purely lytic or early lesions. Scintigraphy does not identify early marrow involvement, although the bone-scan can become positive relatively early in the progression of a metastatic lesion as an osteoblastic reaction occurs[6]. MRI imaging, by contrast, can identify the early marrow oedema and other changes indicative of marrow infiltration by tumour[6]. A visible osteoblastic response may not result from aggressive or purely lytic metastases[63]; this phenomenon commonly occurs in myeloma, and a bone scan is not routinely recommended for evaluating such diseases[63].

Bone scintigraphy is often criticised for having a low specificity[60,64] because false-positive findings on bone scan are common, particularly in the elderly population who have arthritic changes, but are also most at risk of malignancy[65]. The sensitivity of planar bone scintigraphy is dependent on the anatomic site of bony disease, with a sensitivity for spinal and pelvic deposits of only 40%[66]. Unfortunately, the spine is the most common site of metastatic deposits for prostate cancer[67].

Single Photon Emission Computed Tomography (SPECT) images can be obtained, improving lesion detection and localisation[68]. Compared with planar scintigraphy, SPECT has higher sensitivity (87% vs. 74%[68]), specificity (91% vs. 81%[68]) and greater diagnostic accuracy, particularly for vertebral metastases[69]. SPECT imaging can provide useful information in the evaluation of the spine, skull and pelvis, by increasing contrast resolution[5] and allowing appreciation of both the three-dimensional anatomy and the pattern of uptake. It is possible to detect more bone lesions with SPECT compared to planar scintigraphy, and by being able to accurately locate the position, the differentiation between benign and malignant lesions is more accurate[59,65]. One study suggested high-resolution vertebral SPECT produced results comparable with MRI in detecting vertebral metastases, and superior to MRI in the detection of osseous extra-vertebral body metastases[70]. The addition of CT to SPECT has improved the diagnostic accuracy, and improved inter-observer variation significantly[71].

1.1.3.5 Positron Emission Tomography (PET)

1.1.3.5.1 Basic principles of PET

The radiation emission from single photon emitting nuclides has proven, with SPECT, to be a valuable diagnostic tool. The spatial localisation of the tracer accumulation is limited by the random direction in which the emitted gamma-rays can travel. Collimators are used to ensure only gamma-rays travelling near perpendicular to the detectors are admitted into the camera, providing spatial orientation of the originating decay, minimising the impact of scatter. However, this reduces the efficiency of the imaging system. Detection of a pair of gamma photons, produced at 180^{o} to each other from the same point of origin theoretically eliminates the need for a collimator as only 'paired' photons are counted. This is the basis for PET imaging.

The phenomenon of positron annihilation, resulting in the generation of two high energy photons at almost 180^{o} to each other, has revolutionised nuclear functional imaging. Some radionuclides emit positrons during radioactive decay. Positrons are electrons with a positive charge, and have limited penetrating power through tissues. When these positrons encounter a normal negatively charged electron an annihilation reaction occurs during which both particles are converted into energy, emitted as two photons, both with energy of 511keV, travelling out in a straight line from the focus of annihilation at almost 180^{o} degrees to each other. Detectors are placed in a ring around the body to collect the photons for quantification. The detection of this pair of electron almost simultaneously, but on opposite sides of the detection ring, identifies a line through the patient along which the original annihilation reaction occurred. The point of origin is unlikely to be equidistant between the two opposed receptors, and one photon is likely to be detected fractionally before it's pair. More recent scanning technology has enabled further spatial resolution by utilising this 'time of flight' discrepancy to identify where along the line creased by the two opposed photons the annihilation reaction was likely to have occurred. The detection of the emitted radiation from all around the patient enables 3D reconstruction of the images, and all modern PET scanners have an integrated CT scanner - images from both can be superimposed.

Alongside the development of the PET/CT imaging technique has been the synthesis of radio-labelled tracers enabling the targeting of specific tissues, cellular functions or metabolic pathways.

1.1.3.5.2 Qualitative assessment of PET

Qualitative analysis of PET images has a significant role in routine clinical care, both at the level of diagnosis, and to assess disease response. After treatment, no abnormal tissue should persist if a cure is to be achieved, and no abnormal tracer uptake should be seen. In this way, PET imaging, particularly with ^{18}F-FDG, has a significant role in identifying patients with residual disease after therapy. One well established qualitative assessment standardisation is the Deauville Score for recording the chemotherapy response of Hodgkin's Lymphoma[72], using the liver and mediastinum to reference abnormal ^{18}F-FDG tracer uptake, with strong predictive value[73].

1.1.3.5.3 Quantitative assessment of PET

Quantification of PET images can provide additional complementary information to the qualitative analysis, and is potentially more reproducible. Lin et al.[74] showed greater accuracy in prediction of event free survival using SUV quantification of a mid-treatment ^{18}F-FDG PET scan rather than a qualitative analysis. Visual inspection is reader-dependent; a standardised quantification approach would be expected to be more repeatable and reliable. There have been few comparisons of the accuracy from visual inspection to quantification techniques (e.g. Lin et al., 2007[74]).

There are several ways standardised quantifiable data can be extracted from PET images. Table 1-1 summarises the broad methods of image quantification, and compares the merits of each.

Method	Advantage	Disadvantage	Dependency
Visual	Static WB imaging No need blood sampling Short scan times Less reliant on attenuation correction	Subjectivity Single view of dynamic process Dependency on background uptake	Uptake time Patient factors e.g. blood glucose levels for ^{18}F-FDG Partial volume effects
SUV	Static WB imaging Semi-quantitative analysis No need for blood sampling Ease of calculation	Numerous methods of calculation Single view of dynamic process Need for attenuation correction Inaccuracy in detecting small changes	Uptake time Patient factors Body weight Partial volume effects
Kinetic	Dynamic data Quantitative analysis Low dependency on imaging time	Need for the input function Complexity of computation	Partial volume effects Quality of input function

Table 1-1: Methods for assessing PET tracer uptake (adapted from[75])

SUV Analysis

The SUV is the most widely used semi-quantification parameter used in oncology currently, and is the ratio between the radiotracer concentration in a region-of-interest (ROI) and the injected activity, at a single time point, normalised for factors affecting the tracer distribution. This correction factor for ^{18}F-FDG PET is normally patient weight (SUV_{BW}), but can be lean body mass (SUV_{LBM}), or body surface area (SUV_{BSA}). These all have purported benefits in certain scenarios, but require varying levels of patient measurement and practicalities. SUV can be calculated as follows:

$$SUV = \frac{r}{(\frac{a'}{w})}$$

Where *r* is the radioactivity concentration (kBq/ml) within the target lesion, ***a'*** is the decay-corrected amount of injected tracer, and *w* is the weight of the patient (used as a surrogate of volume of distribution for the tracer, where 1ml=1g)[76]

If all the tracer is retained and distributed evenly throughout the patient's tissues, the SUV will be 1. This can be further corrected to the lean body mass of the patient (SUL). SUL is less dependent on body habitus than SUV when compared across the population, and is typically more consistent between patients with ^{18}F-FDG because patients which high body mass have high normal organ SUVs due to the lack of accumulation of ^{18}F-FDG in adipose tissue in the fasting state.

Many factors influence SUV measurement, including patient preparation factors, fasting and blood glucose levels (particularly for ^{18}F-FDG imaging), the time for tracer uptake before the scan is acquired, tracer pharmacokinetics, patient movement, and quantification methodologies. The time of scan acquisition, the image processing and attenuation corrections all have can affect the SUV[77]. The partial volume effect can lead to an underestimation of SUV, and small tumours can appear less avid, and therefore perhaps less aggressive, than they truly are[78-81]. Random errors are introduced into the PET image from image noise. The lower the tracer uptake and lower the resulting measured accumulated activity, the greater the impact from image noise, and the more likely significant random errors will be influence quantification.

It is essential to control and standardise all the scan acquisition and patient preparation procedures, to try and ensure methodological consistencies for comparison of quantification parameters.

Although a change in SUV is typically used to provide a rudimentary quantification of a treatment response[82] there are technical limitations. SUVs do not take consider the concentration of the radiopharmaceutical changing with time, both in terms of plasma concentration and of the kinetics of the tracer in the tissue. These kinetic factors are not necessarily linear, and therefore full dynamic information is required to model the underlying transport processes.

1.1.3.5.4 Semi-Quantification Parameters

SUV_{max} is often reported, being the voxel with the highest SUV within the VOI. This is quick to perform on most clinical PET software platforms, is reproducible, and is reported in most clinical papers[76]. SUV_{max} has become the SUV parameter of choice, not through technical appraisal, but

because of ease of measurement[83]. This is despite the known limitations of SUV_{max}, particularly the high dependence on the statistical quality of the images, and the size of the maximal pixel used for calculation[84]. It is also not representative of the heterogeneity of uptake within a VOI.

SUV_{mean} is a better measure of the population of voxel values within a target, but is impacted by the VOI definition method. The edge of the lesion is unlikely to be clearly defined or reliably delineable; SUV_{mean} is less reproducible than SUV_{max}.

SUV_{peak} is the average SUV usually calculated within a $1cm^3$ sub-volume of a VOI containing the highest average uptake; this VOI may not necessarily contain the single voxel with the highest value in the whole target volume. SUV_{peak} might be less affected by image noise than SUV_{max} because it is calculated from more than a single voxel.[85]. There are many examples of VOI used to calculate SUV_{peak}, varying the VOI size, position and shape[85].

Voxel SUV measurements across a VOI are heterogeneous. There are a number mathematical and statistical ways to describe the distribution of voxel values throughout a volume, describing their image heterogeneity or texture. 1st order heterogeneity parameters are further global descriptors of the distribution histogram and parameters include kurtosis, skewness, energy and entropy. Energy is a measurement of the uniformity of the distribution; a high value suggests the distribution is more uniform. Entropy is a measure of the randomness of the distribution; a high entropy suggests the distribution is spread more widely, and conversely a simple image has a low entropy.

An alternative quantification technique is the direct measurement of focal kinetics of tracers.

1.1.3.5.5 Kinetic Modelling of PET

True quantitative assessment of PET imaging is derived from measurement of the kinetics of the tracer following injection, providing estimates of absolute indices of tracer uptake and retention in the lesions of interest.

To derive kinetic quantification from PET, Van den Hoff described the following requirements[86]:

- A time-activity curve for the blood (i.e. an input function)
- A time-activity curve for the target tissue (i.e. the tissue response)
- A knowledge of the underlying physiology

Hawkins described a compartmental model to identify kinetic measurements necessary to accurately quantify ^{18}F-fluoride uptake, reflecting movement of the tracer between three compartments (plasma, extravascular, and bone). This model is summarised in the figure 1-1.

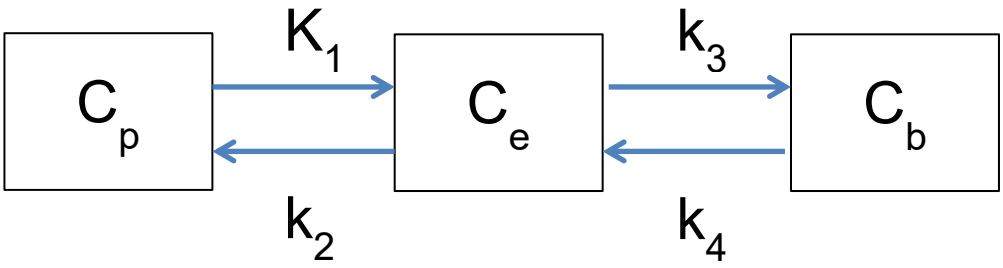


Figure 1-1: 3-comparment model for kinetic measurements, based on[87]. C_p = Plasma compartment; C_e = Extravascular compartment; C_b = Bound to target; K_1 & k_{2-4} = 1st order rate constants describing the potential direction changes between compartments

Using this model, it is possible to calculate the K_i function which represents the net uptake (or influx) of fluoride into bone, reflecting the bone blood flow and the fraction of tracer binding the bone mineral

$$K_i \text{ (ml/min/ml)} = K_1 \times (k_3 / (k_2 + k_3))$$

Such kinetic modelling still requires assumptions: metabolic processes are at a steady state during measurement; there is a homogeneous tracer distribution within each compartment; rate constants are independent of tracer concentration (i.e. display 1st order kinetics)[88].

Assumptions are still needed, but calculation of the Ki function, which represents the net influx of tracer into the bone mineral compartment, has been shown to be a useful quantification measurement independent of changes in tracer kinetics occurring elsewhere in the skeleton, providing a more accurate measurement of local bone metabolism[89]. This methodology has been used to characterise regional bone kinetics in non-malignant disease[90,91] and latterly in bone metastases[89,92].

Using the compartment model, quantification of PET tracer uptake requires knowledge of the activity in the plasma feeding the lesion of interest; this is termed the arterial plasma input function (IF). There are several methods to derive the IF, with continuous or serial arterial sampling being the gold standard[93]. This, however, has obvious logistical (and patient satisfaction) difficulties. It takes considerable time for the activity concentration of venous plasma to match the arterial side and methods of 'arterialised' venous blood sampling have been described, with warming of a periphery causing venous dilatation and venous shunting of arterial blood[94]; inaccuracies with this technique have been identified[77]. Image derived input functions (IDIFs) have been described, and these provide close correlation with serial arterial plasma sampling[95,96], but require fewer blood tests, and the venous blood samples are taken later when the activity (and subsequent exposure to staff) is greatly reduced.

To derive the arterial input function non-invasively a region of interest (ROI) can be placed over an artery, and the activity concentration calibrated against venous blood samples obtained at a later time when the venous and arterial radioisotope activity concentrations equalise[97]. IDIFs have shown good correlation with arterial sampling[95,96].

1.1.3.5.6 Definition of VOI for PET Quantification

A VOI is necessary for quantification. SUV of individual voxels can be measured, but for a whole lesion it is possible to describe the distribution of the voxels, e.g. maximum, mean, standard deviation. The definition of the 'whole lesion' is fraught with difficulties. The partial volume effect influences lesions, particularly smaller lesions. SUV parameters have been shown to be correlated with lesion size, especially SUV_{mean}[98]. Even in larger tumours, if there is heterogeneity of uptake, tracer uptake can be underestimated e.g. an active tumour edge with high uptake, but a less avid necrotic centre. The SUV uptake will be inconsistent across such a tumour, but which voxels relate to the most biologically relevant tissue is unknown.

VOI definition is a vital contributor to the validity and repeatability of semi-quantitative calculation[99-101]. Recognition and delineation are essential tasks for successful segmentation [102], aiming to maximise the inclusion of the pathology being interrogated and minimising the inclusion of the surrounding tissues not under evaluation.

Methods proposed for VOI definition include manual delineation, a fixed volume VOI, an iso-contour based on a percentage of the VOI SUV_{max}, and more complicated computation and statistical approaches adapting the segmentation of the VOI based on the target:background ratio or algorithms to analyse voxel relationships. Attempts have been made to standardise methodology, including development of the PERCIST criteria[83], but no consensus has been reached. Table 1-2 summarises the most commonly used VOI methods.

Technique	Summary	Benefits	Limitations
Manual	Visual interpretation and selection of ROI	Simple	Time consuming Variations with window settings Inter- and intra-observer variation[103-105]
Fixed dimensional ROI	Manual placement of fixed volume over area of interest which may include SUV_{max}	Simple	Will not provide whole lesion or reproducible metrics
Thresholding	Irregular iso-contour based on a fixed percentage of SUV_{max}, or fixed absolute SUV level, or based on background threshold value	Relatively simple Quick Most frequently used	Selection of optimal threshold Sensitive to partial volume effects, tumour heterogeneity and motion
Adaptive approaches	Mathematical algorithms utilising target and background signal intensity	Boundary continuity Relatively efficient Better for heterogeneous distributions	Sensitive to image noise
Learning methods	Pattern recognition; complex algorithms comparing voxel values to surrounding voxels, e.g. clustering	More advanced voxel analysis	Computational complexity
Stochastic models	Statistical differences analysed between target uptake and surrounding tissues	Possibly less susceptible to image noise	Computational complexity

Table 1-2: Summary of PET image VOI methods, adapted from Zaidi et al[106].

Applying segmentation threshold levels to a target lesion is now a frequent approach to improve repeatability of PET quantification. These techniques are usually based on the SUV_{max} of the target and may be affected by image noise. One key advantage is the near universal ability for PET centres to be able to perform this without the requirement for additional complex software. The choice of threshold level impacts the calculated lesion parameters significantly[84]. Erdi et al. identified a threshold of around 40% of SUV_{max} on ^{18}F-FDG PET images best represented the volume of disease as identified on CT[107]; others used this threshold level for VOI definition for delineation of cervical cancer[108], non-small cell lung cancer[109] and head and neck cancer[110]. Whether this accurately delineates the full tumour lesion is unclear, and histological correlation is lacking. The optimal threshold is likely to be highly dependent on the tumour type, the radiotracer and the tumour volume[111,112]. Nestle et al. demonstrated up to a 41% difference in the VOI generated by a variety of delineation methods, including manual, threshold methods (absolute and relative to SUV_{max}), and adaptive[101], when using ^{18}F-FDG PET to delineate lung tumours.

With the advent of increased computational processing power has come the application of more complex adaptive VOI delineation. Several different techniques have been described, with the proffered advantage of being less susceptible to image noise and contrast variations affecting more conventional threshold methods[113]. Hatt, in 2009 reported a Bayesian statistical image segmentation method termed FLAB (fuzzy local adaptive Bayesian)[114]. This algorithm was designed to improve upon earlier statistical approaches to image segmentation. They compared their model with other algorithm methods and concluded FLAB was superior for lesions less than 2cm compared to other algorithms, and they reported errors of under 10% for non-spherical lesions with inhomogeneous distributions of activity, with high levels of reproducibility with variation less than 4%[113].

1.1.3.5.7 ^{18}F-Fluoride PET

Uptake of ^{18}F-fluoride PET tracers is similar to ^{99m}Tc-MDP standard bone scintigraphy, with absorption by hydroxyapatite. The uptake is more rapid because the ^{18}F fluoride ions are not bound to plasma proteins[115]. Fluoride ions diffuse from the intravascular space into the extracellular fluid compartment, and bind onto the bone surface through exchange with the hydroxyl groups in the hydroxyapatite crystals of the bone, resulting in fluoroapatite[90]. The uptake is related to the regional blood flow and to osteoblastic activity; the tracer accumulates in areas of osteoblastic activity[39]. The increased capillary permeability and increased local blood flow due to metastatic deposits in bone consequently mean greater uptake in bony metastases compared to normal bone[115]. ^{18}F-Fluoride PET provides the opportunity to quantify features of the local environment to tumour deposits rather than the tumour cell directly, assuming surrogacy of the bone environment and tumour cell activity.

^{18}F-Fluoride PET/CT has a very high sensitivity and specificity (100% and 97% respectively[61]) for the detection of bone metastases. Bone-specific imaging with ^{18}F-Fluoride PET/CT has shown increased diagnostic accuracy when compared with technetium based standard scintigraphy and SPECT[66], and has the benefit of aiding differentiating between benign and malignant bone lesions[116].

1.2 Assessing treatment response of bone metastases

Accurately and timely response assessment is a clinical priority. Measuring soft tissue tumours in situ is fraught with difficulties. Even physical measurement of tumours by experienced clinicians is hampered by intra- and inter-observer variability[117].

It has been vital in oncology to strive towards a standardisation of response definitions to allow universal understanding, and to allow homogeneity for inter- and intra-patient comparison, particularly important for clinical trial outcomes. The standardisation of response criteria has passed through several variations and iterations, and there are specific points to consider depending on the imaging modality being reviewed. Qualitative assessment of imaging is still the mainstay of oncological response assessments, alongside tumour size measurements, and is the only current standard method for assessment of bone metastases.

The World Health Organisation (WHO) attempted in the 1980s to standardise measurement of soft tissue tumour deposits using plain films and bone scans. These initial criteria were based on bi-dimensional measurements[118]. These definitions did not create the standardisation desired, with up to 40% major disagreements identified[119]. A response in bone lesions was to be determined by a reduction in size or re-calcification of lytic lesions, a decreased density of blastic lesions, and stable disease could not be concluded before 8 weeks due to the slow response of bone lesions[118,119]. The International Union against Cancer (UICC) criteria for categorising treatment response in breast cancer, published in 1977, term bone metastases as ‘evaluable, but non-measurable’. These guidelines suggest objective evaluation of bone lesions using X-rays and scintigraphy, including assessment of healing sclerosis of lytic lesion, whether new lesions

develop[120] This was clearly imperfect, and is not relevant to predominantly sclerotic bone metastases, including those from prostate cancer.

The RECIST criteria followed (version 1.0 in 2000[121], version 1.1 in 2009[122]), and uses the sum of the longest diameters. Response is identified by a change in size of the target lesions, with PR defined as ≥30% decrease (but not a complete response) and PD as ≥20% increase[122]. These criteria still have limitations. RECIST criteria depend predominantly on the measurement of solid tumours. Bone metastases are not readily measurable.

The inability to include bony metastases in the RECIST criteria led to the development of a bone-specific response criteria which took into account the significant advances in imaging technology, including findings on plain X-ray, CT and conventional MRI imaging, and skeletal scintigraphy; the MD Anderson (MDA) criteria[34] includes qualitative assessments of bone metastases. PET and SPECT were not included in the criteria because there was, at the time, no evidence to support their application in response assessment of bone metastases[34].

The development of functional imaging methods to image and quantify the underlying tumour cell metabolism characteristics raises the need for alternative definitions of response: metabolic response criteria. This is particularly relevant with the development of new anti-cancer therapies aiming to have a cytostatic effect, rather than causing tumour cell death and demonstrable tumour shrinkage; lack of progression of tumours may be associated with improved outcomes, even when a size response has not been identified[83].

Although there has been widespread integration of functional imaging techniques into clinical practice, and the availability and cost has fallen significantly, there has been much less integration into clinical trial design, principally due to the variation in technique between centres, and a lack of consensus on the appropriate treatment response metrics and definitions, and how PET might be used as a surrogate endpoint for assessing clinical efficacy of new therapies[75,83].

The EORTC group began the work to bring uniformity to PET response assessment, producing their first recommendations in 1999[123] (see Table 1-3).

Progressive Metabolic Disease	An increase in ^{18}F-FDG >25% within tumour region defined on baseline scan, or a visible increase (>20%) in longest dimension, or the appearance of new lesions
Stable Metabolic Disease	An increase in tumour SUV of<25% or less than 15% with no visible increase in extent of uptake (i.e.<20% increase in longest dimension)
Partial Metabolic Response	A reduction of minimum 15-25% in tumour ^{18}F-FDG SUV after one cycle of chemotherapy and greater than 25% after one treatment cycle (25% was a useful empirical cut off, but there was no reproducibility assessment at this stage). A reduction in the extent of tumour uptake was not a requirement.
Complete Metabolic Response	No significant uptake of tracer

Table 1-3: EORTC recommendations for defining treatment responses using ^{18}F-FDG PET (adapted from[123]).

Since publication of this initial guidance there have been attempts to standardise the PET scan protocols and analysis with an aim of enabling integration of PET scanning into clinical trials.

Shankar et al. reported recommendations of a workshop panel where the main factors influencing the variation in PET scan technique and quantification were considered[75]. The recommendations included patient preparation standardisations and advice on scan timings. Note was made of the importance of attenuation correction methods, ROI/VOI definition methods, tracer dose, and quantification parameters to be used. The lack of robust comparative data limited the recommendations provided by this group. SUV_{max} and SUV_{peak} were felt to be the most robust measurements, but the details around ROI definition, attenuation correction and tracer dose could not be resolved[75].

The Netherlands protocol "for standardisation and quantification of ^{18}F-FDG whole-body PET studies in multi-centre trials"[124] addressed many of these issues, aiming to provide a standard of PET scanning enabling consistency between centres. This group addressed patient preparation, methods to match scan statistics, image reconstruction issues, and they devised a multicentre quality control phantom. They also made recommendations about VOI definition and SUV calculations[124]. Definition of response using SUVs are, however, a further level of complexity still lacking consensus.

The PERCIST criteria[83] were a proposed RECIST style approach to standardising definition of response using ^{18}F-FDG PET, developing metabolic response definitions reliant on standardisation of the scan acquisition and quantification variables known to be influential. The key recommendations are summarised briefly in Table 1-5:

Variable		Recommendation
SUV normalisation		SUL
Target lesions		Hottest single tumour lesion identified (highest SUL of maximal 1.2cm ROI in tumour). Up to 5 measurable lesions ("typically the hottest which are typically the largest")
Quantification parameters		SUL_{peak}; SUL_{max} Also record SUL_{mean} using 50% and 70% SUL_{peak} iso-contours; TLG
Response Categorisation	**Complete Metabolic Response (CMR)**	Complete resolution of ^{18}F-FDG uptake within target lesion
	Partial Metabolic Response (PMR)	At least 30% reduction SUL_{peak} and no new lesions
	Stable Metabolic Response (SMR)	Not CMR, PMR or PMD.
	Progressive Metabolic Disease (PMD)	>30% increase in ^{18}F-FDG SUL_{peak} OR – visible increase in extent of tracer uptake OR – new lesions

Table 1-4: PERCIST key recommendations (Adapted from[83]).

The aim of these recommendations was to provide consistency for using ^{18}F-FDG PET in clinical trials. The guidelines do not aim to identify the most discriminating quantification parameters, and they do not broach the myriad of PET tracers available in clinical practice, including ^{18}F-Fluoride PET, and bone metastases are not specifically identified. These recommendations are clearly not applicable to other functional imaging methods such as DW-MRI.

1.3 Imaging treatment response

1.3.1 Response assessment with scintigraphy

Assessment of a treatment response with scintigraphy is seen by a qualitative change in the tracer uptake. This has limitations. Growth of lytic lesions, demonstrating predominantly osteoclastic activity, does not result in ^{99m}Tc MDP tracer uptake[125]. The flare phenomenon is well characterised, seen within 4-12 weeks from treatment onset[126-128], whereby new bone formation by osteoblastic activity in healing bone metastases following a treatment response results in increased tracer uptake, which could be misinterpreted as progressive disease(PD)[39]. In a group of breast cancer patients, a flare response was noted in 35% at a mean time of 3.3 months following start of treatment[129]. It can take 6 months for a flare response to reduce[130], with others reporting persistence at 2 years[131]. For this reason the Prostate Cancer Working Group (PCWG2) guidance on assessing a treatment response in bone metastases from prostate cancer recommend a bone scan not before 12 weeks after initiation of therapy, and, if new lesions are identified, a confirmatory scan should be repeated 6 or more weeks later (i.e. at 18 weeks), delaying confirmation of PD[132].

1.3.2 Imaging Response Assessment with CT

CT imaging can be used to monitor the sclerotic change in bone metastases following a response to treatment. Quantification using Hounsfield units has tracked the change in reactive sclerosis within metastatic deposits after treatment with bisphosphonates[133]. Bone metastases can result in associated soft tissue extension, which enables RECIST size criteria of response to be applied. New lesions can be identified, although the appearance of a sclerotic lesion with treatment can often suggest a treatment response in a previously unseen small lytic deposit.

1.3.3 Imaging Response Assessment with Quantification of PET

The feasibility of detecting metabolic changes with ^{18}F-FDG PET in extra-cranial tumours following treatment was initially shown when monitoring neoadjuvant chemotherapy for breast cancer patients[134], showing a SUV reduction of up to 50% in primary tumours after 3 months in those who responded; responses were noted in SUV measurement after only 8 days of treatment. Changes in ^{18}F-FDG SUV are now widely used for monitoring treatment responses for a wide variety of tumours, but there is still no consensus on how to standardise and define metabolic responses.

Bone metastases have been an obvious target for research given the poor value of monitoring morphological changes to therapy. Stafford et al. in 2002 reported a 24-patient study of serial ^{18}F-FDG PET scans for women with bone metastases from breast cancer; changes in SUV correlated with the clinical response assessment[135]. A 2006 study of spinal metastases treated with stereotactic radiosurgery showed changes in ^{18}F-FDG uptake 1 and 6 months following treatment correlated with the clinical response[136].

A later study of breast cancer bone metastases with 25 patients clearly demonstrated sequential ^{18}F-FDG scans showing changes in the tumour activity in bone metastases following treatment,

whereas morphological changes were inconsistent[137]. Breast cancer bone metastases were monitored with ^{18}F-FDG PET before and after treatment by Tateishi et al., reporting in 2008. A significant correlation was identified between increases in CT attenuation with decreases of ^{18}F-FDG SUV and TLG[138].

Other tracers have also been interrogated. Lindholm et al. reported in 2009 a study of 13 breast cancer patients with bone metastases, imaged with ^{11}C-methionine PET, concluding a significant fall in SUV in responding sites of disease[139]. More recently, Cook et al. have demonstrated sequential ^{18}F-Fluoride measurements of prostate cancer bone metastases before and after therapy show correlations between change in SUV and PSA[140].

1.3.4 Imaging Response with DW-MRI

Most studies demonstrate an increase in tumour ADC values following successful therapy[41]. Following successful therapy, tumour cell death and vascular change in response can precede morphological changes, raising the possibility of using DW-MRI as a response biomarker. An increase in ADC following successful treatment has been identified as early as 6-8 days after treatment[141], preceding measurable tumour regression, and has been histologically attributed to an increase of extracellular space[141]. There is some uncertainty of the underlying mechanism associated with changes in diffusion, but an increase in ADC after therapy for several tumour types has been demonstrated, including in breast tissue xenografts after only 2 cycles of chemotherapy[142]. A National Cancer Institute sponsored consensus meeting in 2008 from the Society of Magnetic Resonance in Medicine set out recommendations to try and achieve consistency in imaging techniques, analysis and reporting for studies examining response assessment with DW-MRI[41], with the overall aim of establishing DW-MRI as an imaging biomarker.

A response to radiotherapy of spinal lesions was suggested by Byun et al. in 2002[58] with repeat imaging 6 months after treatment. The authors however reported an increase in ADC in 23 patients who had a symptomatic benefit from radiotherapy to the target lesion. T1 and T2 MRI imaging of the same lesions did not detect qualitative changes.

Further studies demonstrated a rapid increase in ADC following successful therapy, including in liver lesions from breast cancer treated with chemotherapy (the investigators could predict a response with ADC just 4 days after therapy[143], and also in brain tumours treated with radiotherapy[144]. A low ADC, it is suggested, identifies tumour viability, compared with a relatively high ADC secondary to necrosis which would confer radiotherapy resistance[144]. Cui et al.[145] monitored ADC changes in patients with hepatic metastases from colorectal or gastric primary tumours. Lesion with a lower pre-therapy ADC were more likely to show a treatment response, with ADC noted at 3-7 days. Significant correlations were identified between the overall reduction in size of the lesions and both the pre-treatment ADC and early ADC changes[145]. Although a different histology, Thoeny et al. in 2005 demonstrated the link between ADC and necrosis. They measured the ADC of rhabdomyosarcomas in rats before and after therapy, and compared the imaging quantification with histology. They showed an increase in ADC two days after therapy

correlated with histological necrosis, and a subsequent decrease in ADC nine days after treatment corresponded with tumour regrowth[146].

Not all studies have demonstrated value from measuring ADC before treatment. Manton et al. were unable to predict response to neoadjuvant chemotherapy for primary breast cancers using baseline ADC measurement[147].

Changes in ADC were noted in primary bone tumours after treatment by Yakushiji et al.[148], showing a greater change in the ADC in the tumours with more necrosis before treatment. There was, however, no correlation attempted between change in ADC and clinical outcome. In another small study of histologically proven osteosarcoma patients Uhl et al. demonstrated higher ADC values in necrotic areas compared to viable tumour[149] following chemotherapy treatment.

Lee et al. used a preclinical model of metastatic prostate cancer to investigate the feasibility of using MR diffusion quantification to assess the change in lesions after docetaxel chemotherapy treatment[150]. Significant increases in ADC were detected seven days after therapy, and histological correlation showed previously dense cellular morphology had changed to a bone lesion with more heterogeneous density[150]. The same team translated this pre-clinical experience to a real-life clinical setting, imaging a patient with bone metastases from prostate cancer with DW-MRI before initiation of anti-androgen treatment (not chemotherapy), and again at 2 weeks and 8 weeks after treatment started. Clinically the patient responded to therapy, and after 8 weeks there was an increase in ADC measured, correlating with a decreased PSA suggesting a response, but no demonstrable change in size of the bony lesions measured[151].

The first clinical series to demonstrate ADC changes in bone metastases of patients undergoing treatment involved 26 patients with prostate cancer bone metastases who had lumbar spine and pelvis DW-MRI imaging before chemotherapy and again at 12 weeks following treatment[152]. Using standard response criteria, 8 patients were identified as responders and 15 progressors. 100 lesions were analysed in total, but there was no significant difference identified in pre-treatment ADC between responders and progressors. However, in those patients who responded to treatment there was a significant increase in ADC after treatment[152]. Interestingly, there was also a significant increase in ADC of the lesions measured in those patients who progressed on treatment, and there was heterogeneity of ADC response within both responders and non-responders. The authors suggested the magnitude of ADC rise was greater in responders, with a sensitivity of 75% and specificity of 66.6% for identifying response if a minimum of 25% increase in ADC was used as a cut-off. The attribute the heterogeneity of ADC change to changes within the composition of the bone marrow as well as histopathological changes in the tumour.

Reischauer et al.[153] prospectively monitored 9 treatment-naive men with bone metastases from prostate cancer with DW-MRI imaging before treatment and again at 1, 2 and 3 months after anti androgen therapy was started. They reported significant increase of ADC_{mean} of metastases, even after 1 month[153]. Clinically, patients responded, with PSA dropping by more than 90%[153]. These authors also noted a heterogeneous response on functional diffusion maps which showed some voxels had a fall in ADC, and more voxels demonstrated this at 3 months after therapy compared to the 1-month imaging.

Most recently, a group in London reported on their development of a semiautomatic segmentation method of whole-body diffusion weighted MRI in an attempt to derive indices of tumour burden[154]. Whole-body diffusion-weighted imaging was performed on 11 patients in total, 4 female patients with metastatic breast cancer and 7 male patients with metastatic prostate cancer. Imaging was completed before treatment and then again at between 12 to 38 weeks after therapy was started. Response was assessed clinically based on other imaging, clinical assessment and measured tumour markers. The authors developed a semi-automatic segmentation algorithm to select areas of bony disease. Those patients with a treatment response had a significantly larger increase in median global ADC compared to non-responders, and the non-responders had a significantly larger increase in total diffusion volume (i.e. volume of disease segmented)[154]. Non-responders had a small decrease in median gADC after treatment.

1.4 Novel imaging approaches for appraisal of bone metastases

1.4.1 Whole-Body DW-MRI

A limiting factor in the use of DW-MRI imaging for spinal assessment was the size of the target organ; technical advances have been necessary to enable efficient scanning of larger areas to include most of the axial skeleton. One of the earliest approaches to whole-body MRI imaging was reported by Steinborn et al. in 1999. They evaluated the potential diagnostic benefits of a whole-body bone-marrow protocol (fast T1-weighted and STIR images) for the detection of bone metastases. They were able to identify 216 lesions in 18 patients; standard bone scans of the same patients only identified 159 lesions[155].

Takahara et al.[156] were among the first to describe a technique for a more rapid whole-body three-dimensional diffusion weighted image with background body signal suppression[156]. Their technique allowed for free breathing, and resulted in a much faster whole-body technique. Such a technique was used for assessment of haematopoiesis in the bone marrow in the same year[157].

Nakanishi et al,[158] compared a whole-body DW-MRI technique with standard bone scintigraphy for the detection of bone metastases. The authors reported a sensitivity of 96% and positive predictive value of 98% when DW-MRI was used in conjunction with T1 and STIR MRI images[158]. For the same patients, they reported a sensitivity of 96% for skeletal scintigraphy. This suggested DW-MRI imaging of the skeleton was certainly capable of detecting metastatic bone disease successfully, but did not suggest superiority to current standard imaging. This was also concluded by Xu et al. the following year using a 3T MRI scanner[159].

The benefit of adding whole-body diffusion weighted MRI to standard MR was considered by Takenaka et al.[160]. This group also compared DW-MRI with ^{18}F-FDG-PET and standard scintigraphy for detecting bone metastases in 115 consecutive non-small cell lung cancer (NSCLC), with a 12-month clinical follow up being used as a standard of reference. They concluded the specificity and accuracy of whole-body MR with DWI (95.5% and 96.1% respectively) were significantly higher than scintigraphy (reported by the group as 95.5% and 95.0%); the improvement in specificity was small, but statistically significant. This benefit came from combining whole-body MR with DWI. They reported integrated a sensitivity of 97% and

specificity of 95.4% using ^{18}F-FDG PET/CT. The authors conclude whole-body DW-MRI is as accurate as scintigraphy and/or ^{18}F-FDG PET/CT, with the benefit of reduced radiation exposure[160]. Similar findings have been found by several other groups[50,51,161,162].

Gutzeit et al.[162] compared DW-MRI and skeletal scintigraphy for the detection of bone metastases from prostate cancer and breast cancer, prospectively scanning 36 patients in total. They analysed the b=1000s/mm^2 images, and separated identified lesions into three categories of pathological suspicion. They concluded a sensitivity for detecting metastatic lesions at 97% (82%-100%) if there were more than ten lesions, but only 58% (29%-84%) if there were fewer than 5 lesions (scintigraphy 48% (31%-66%) and 67% (35%-89%), respectively). The authors concluded non-superiority of DW-MRI over skeletal scintigraphy. Patients with a higher burden of bony metastatic disease DW-MRI seemed to perform better than a bone scan.

In 2011 Wu et al. performed a meta-analysis questioning the diagnostic accuracy of whole-body MRI for detection of bone metastases[163]. This did include a range of MRI techniques, but a subgroup analysis was performed for those who had diffusion-weighted imaging (DWI). The authors concluded DWI was sensitive, but led to reduced specificity (pooled sensitivity for DW-MRI imaging 87.7% (76.3%-94.9%), pooled specificity 86.1% (79.2%-91.4%)) compared to whole-body MRI without DWI.

Elber et al. evaluated methodological aspects of whole-body DWI for staging prostate cancer bone metastases[164]. They compared the imaging of the identified bone metastases with the level of sclerosis identified on corresponding CT scans, concluding superior contrast-to-noise for non-sclerotic bone metastases compared to the heavily sclerotic lesions. They also showed densely sclerotic lesions were best visualised on T1-weighted MRI imaging, and were noted to have more restriction of diffusion[164].

1.4.2 Image Heterogeneity

The voxels than comprise the three-dimensional imaging of all modern medical imaging techniques can be considered as individual packets of information, and offers another possible route of image quantification. Individual voxel information is routinely used when quantifying medical imaging. The SUV_{max} of PET studies is the most widely used parameter, largely due to the ease of acquisition – it is simply recording the value of the voxel with the highest calculated SUV within the target VOI. However, the voxel intensities across a tumour or normal tissue VOI are not consistent, and it is not clear which part of a tumour VOI is the most clinically relevant. The histogram of voxel information grained from VOI image analysis can be described in several ways; the mean and median of the distributions are often reported during clinical studies of PET imaging. DW-MRI imaging favours these as measures of water diffusion within a VOI. These parameters however discard most of the other histogram data, and this data about the image heterogeneity may have significant clinical relevance.

Image heterogeneity analysis has been developed over the past few decades for wide ranging applications, from NASA examining satellite images[165], through the study of sea ice patterns[166] and the commercial classification of cashew nut kernels[167], to the current investigations of

possible applications in oncology, there is significant interest in unlocking potential patterns and data from images.

Each individual voxel intensity in medical imaging is the result of numerous interacting factors, and there are many ways the voxel intensity distributions can be described and compared, and possibly used to reveal clinically relevant information. The 1st order heterogeneity or texture features are the descriptors of the voxel intensity frequency-distribution histogram, including the mean and median parameters. The histogram can be interrogated further; the skewness and kurtosis describe the shape of the distribution curve in more detail, where kurtosis is a descriptor the sharpness of the peak of the distribution, and skewness is a measure of the symmetry of the distribution. Entropy of an image is used to describe the randomness of an image and energy is a measure of voxel uniformity.

Computational power has enabled image analysis at more complex levels. The voxel to voxel relationships can be interrogated, measuring how often voxels of differing (or similar) intensities are related to each other, and in what configuration. In addition, more complicated fractal pattern analysis and Bayesian statistical models can be used to categorise imaging features even further.

This all has potential clinical utility, possibly allowing interrogation and identification of specific tissue architectures and behaviours. As early as the 1980s there was interest in the fractal nature of vasculature[168]. Texture analysis has been used to identify neovascularisation[169] and differentiate between benign and malignant tissue[170].

Texture analysis with has been used to predict cancer treatment responses for a number of different soft tissue malignancies and with different imaging modalities, including CT images[171] and MRI scans[172]. Texture analysis of soft tissue malignancies have also shown prognostic potential, including quantification of CT scans[173-177], and PET images[178-180]. The predictive and prognostic value of texture analysis of bone metastases has not been reported in the literature.

1.5 Summary

Bone metastases are common sequelae of common malignancies, and result in significant morbidity. Patients with prostate and breast cancers, even with bone metastases, have prolonged survival, and treatments to further prolong survival and mitigate symptoms from bone metastases are vital. There are now many treatments shown to have excellent clinical value, but assessment of treatment response in patients with only bone disease remains a clinical problem. For this reason, bone metastases are excluded from clinical trials, to the detriment of the breast and prostate cancer patients with bone predominant disease. There is an urgent requirement for the development of accurate imaging biomarkers for assessment of treatment response in bone metastases.

1.6 Scope and objectives

This thesis reports the initial data from 3 prospective clinical trials designed to interrogate, amongst other functional imaging techniques, ^{18}F-Fluoride PET and DW-MRI for the appraisal of bone metastases from breast and prostate cancer, the quantification of treatment response, and correlation between the quantification parameters and clinical markers of disease response.

These studies are, at the time of submission, still recruiting patients; it is beyond the scope of this thesis to report on the objectives of these three clinical studies.

The main objective of this thesis is to identify the best methods for the final analysis of the three clinical studies, ensuring accurate and clinically relevant quantification of the images to maximise any potential clinical utility.

The aims of this thesis are:

^{18}F-Fluoride PET:

- To investigate iso-contour delineation of volumes of interest for quantification
- To investigate quantification parameters with a per-lesion analysis for prediction of the clinical treatment response assessment, and evaluate the quantification parameters in terms of OS and PFS
- To investigate quantification parameters with a per-patient analysis for prediction of the clinical treatment response assessment, and evaluate the quantification parameters in terms of OS and PFS
- To investigate the impact of different VOI methodologies on the quantification parameters
- To develop a quantification method for the whole-skeletal metastatic burden, and assess for predictive and prognostic clinical utility

DW-MRI:

- To investigate quantification parameters with a per-lesion analysis for prediction of the clinical treatment response assessment, and evaluate the quantification parameters in terms of OS and PFS

- To investigate quantification parameters with a per-patient analysis for prediction of the clinical treatment response assessment, and evaluate the quantification parameters in terms of OS and PFS
- To develop a quantification method for the whole-skeletal metastatic burden, and assess for predictive and prognostic clinical utility

1.7 Hypothesis

Quantification of the functional imaging techniques will provide more accurate and timely assessment of a treatment response of bone metastases compared to current clinical and radiological standards.

1.8 Overview of thesis

The three prospective clinical studies to which the patients in this thesis were recruited are summarised in the Trial Design (Appendix 1).

The specific methods applied for development of this thesis are described in the Methods and Patients (Chapter 2) chapter. Following explanation of general methods applicable to all studies in this thesis follows study-specific methods. The subsequent technical chapters interrogate ^{18}F-Fluoride PET and DW-MRI, examining VOI methodologies, and the predictive and prognostic value using per-lesion analyses, per-patient analyses and whole-skeletal metastatic burden analyses. These will be detailed in the Methods and Patients chapter.

Chapter 2 Methods and Patients

2.1 Introduction

Three prospective clinical studies have been designed to interrogate functional imaging methods for the assessment of a treatment responses in bone metastases from breast and prostate cancer; details can be reviewed in Appendix 1.

In preparation for full recruitment and final analysis of these studies, the data from the initial patients has been used to prepare and interrogate quantification methodologies. There are uncertainties about which quantification parameters might be most predictive and prognostic, how to define the quantification volumes of interest (VOIs), and how to define changes between scans.

At the time of initial data analysis for this thesis, patient recruitment to the three studies had supplied sufficient data for analysis of the ^{18}F-Fluoride PET scans and DW-MRI scans.

2.2 Overall Methods

2.2.1 Patient Selection

For all studies, patients were identified with bone predominant metastatic disease, and were recruited before embarking on a change in treatment for radiologically confirmed progressive bone disease.

Full details of the clinical studies can be found in the Trials Design (Appendix 1), but have been briefly summarised in Table 2-1.

FAB-P	**Tumour Group**	Prostate Cancer
	Treatment	Docetaxel chemotherapy
	Imaging	^{18}F-Fluoride PET/CT, ^{11}C-Choline PET/CT, WB + DW-MRI
	Scan Schedule	Baseline, 8 weeks from initiation of therapy (option of 12 week ^{18}F-Fluoride PET/CT scan if possible flare noted at 8 weeks)
	End Point	Clinical assessment of treatment response up to 24 weeks from treatment initiation
FAB-B	**Tumour Group**	Breast cancer
	Treatment	Hormonal therapy +/- bisphosphonate/denosumab
	Imaging	^{18}F-Fluoride PET/CT, ^{18}F-FDG PET/CT, WB + DW-MRI
	Scan Schedule	Baseline, 8 weeks from initiation of therapy (option of 12 week ^{18}F-Fluoride PET/CT scan if flare noted at 8 weeks)
	End Point	Clinical assessment of treatment response up to 24 weeks from treatment initiation
FAB-IE	**Tumour Group**	Prostate cancer
	Treatment	Hormonal therapy/chemotherapy
	Imaging	^{99m}Tc-Maraciclatide Scintigraphy/SPECT/CT, WB + DW-MRI, DCI-MRI
	Scan Schedule	Baseline and 12 weeks from initiation of therapy
	End Point	Clinical assessment of treatment response up to 24 weeks from treatment initiation

Table 2-1: Summary of FAB-P, FAB-B and FAB-IE

These studies have Ethics and ARSAC Committee approvals (FAB-P - 12/LO/0830; RPC 261/3186/28833, FAB-B 12/LO/1801; RPC 261/3186/29516, and FAB-IE 13/LO/0041 RPC 54/3186/29826 respectively).

At the time of this initial data analysis, 9 patients had been recruited into FAB-P, 5 into FAB-B, and 6 into FAB-IE.

All patients met the inclusion criteria for each study: patients (aged 18 or over) with confirmed skeletal skeletal-predominant metastatic breast/prostate cancer who are treatment naïve, or with PD embarking on new systemic anti-cancer treatment; willingness to comply with study visits and tests

Patients with concomitant uncontrolled medical conditions were excluded, along with those patients who had contraindications to MR imaging, those patients likely to require palliative radiotherapy between study scans, those patients likely to require G-CSF (excluded because of the impact on bone marrow activity which can influence DW-MRI imaging quantification).

Written informed consent was obtained from the patients.

2.2.2 Imaging Protocols

2.2.2.1 ^{18}F-Fluoride PET/CT Procedure

The PET scans were carried out in the PET Centre at St Thomas' Hospital using a Discovery VCT (until Oct 2013) and then a Discovery 710 (both GE Healthcare). A target administered activity of 225MBq was injected (Min 150MBq, Max 250MBq).

PET scan acquisition was started 60 minutes after injection of the tracer. Venous blood samples were taken at 55 minutes after injection, and at the end of the PET-CT acquisition (90 minutes after injection) from the opposite arm to the injected activity. The exact times of sampling were recorded. 0.2ml of whole blood and 0.2ml of serum underwent radiation count sampling, and the time of measurement recorded to correct for decay. The ^{18}F-fluoride PET half body was acquired from the lower limbs towards the head in 6-7 frames of 4 minutes each. PET images were reconstructed using ordered-subset expectation maximisation (2 iterations, 20 subsets). A CT scan for attenuation correction was acquired immediately preceding PET acquisition, with a slice thickness of 3.27 mm and a voxel size of 4.7 mm, and PET attenuation correction of the PET was performed (120kVp, Auto mA).

2.2.2.2 DW-MRI Imaging Protocol

All studies were performed on a 1.5T MRI scanner (Magnetom Aera, Siemens, Erlangen, Germany) with an 18-channel external phased-array surface coil (Siemens, Erlangen, Germany), with an echo-planar spin-echo technique (free breathing, STIR fat suppression), scanning from skull-base to mid-thigh). Table 2-2 summaries the MRI imaging protocol.

	T1 Dixon Axial	T2 Haste Axial	Diffusion Axial	T2 Spine Sagittal	T1 Spine Sagittal
TR(ms)	4.77	336	12990	3500	630
TE(ms)	2.39	96	67	97	9.8
FOV	430	430	430	425	425
Matrix	244*320	128p*256	208p*256	307*384	307*384
B-value	n/a	n/a	0,800	n/a	n/a
NEX	1	1	2/5	2	2
ST(mm)	5	5	5	4	4

Table 2-2: Summary MRI protocol used for FAB-P, FAB-IE and FAB-B

2.2.3 Image Quantification

Image quantification specifics, parameter definitions and VOI definitions will be discussed in the separate sections of this chapter which form the methods for the individual results chapters reported in this thesis.

2.2.3.1 ^{18}F-Fluoride PET Quantification Parameters for Analysis

The most commonly parameters are those describing the SUV voxel population within the VOI, namely the maximal voxel value (SUV_{max}), and the mean of the distributions (SUV_{mean}). SUV_{max} is the most widely used clinical parameter, and is operator-independent, although as a measure of a single VOI voxel it may be affected by noise[85]. SUV_{mean} and SUV_{max} have been shown to be reproducible[181], but the use of average measurements will be dependent on the definition of the

VOI, whereas the SUV_{max} will be unaffected[182]. SUV_{peak} has been developed as an alternative less susceptible to image noise, and has been shown to be reproducible[85].

True quantification parameters, measuring the pharmacokinetic movements of the PET radioisotope between the physiological compartments. This compartmental modelling, frequently referred to as kinetic modelling, is considered the most accurate way to analyse PET data[183]. Kinetic modelling has required dynamic scanning, taking multiple frames from the point of injection alongside arterial blood sampling to measure the time course of the tracer concentration in the plasma, used to define the input function for kinetic modelling.

To derive the arterial input function non-invasively a region of interest (ROI) can be placed over an artery, and the activity concentration calibrated against venous blood samples obtained at a later time when the venous and arterial radioisotope activity concentrations equalise[97]. IDIFs have shown good correlation with arterial sampling[95,96]. Siddique et al. described in 2012 a method to estimate regional bone metabolism from whole-body ^{18}F-Fluoride PET static images, and it is this methodology used in this thesis analysis to calculate the regional bone plasma clearance function Ki[184]. This method uses venous blood sampling to estimate the arterial input function by rescaling for injected activity a population residual curve of the ^{18}F-Fluoride bolus peak and exponential clearance from the plasma (obtained from a study of 10 patients by Cook et al.[97]), and combining this to the individual patient's terminal exponential curve measured with the venous blood sampling[184]. This technique has been shown to have close correlation to invasive kinetic modelling with a mean ratio of 1.006 between the static Ki method and the gold-standard dynamic Patlak Ki method, 30 minutes after injection, decreasing to 0.965 at 60 minutes[184], and has a clear methodological advantage of being much less invasive.

Volumetric parameters have been used to describe the metabolic tumour volume (MTV) of tumours, which is the volume of the defined VOI. This parameter, from ^{18}F-FDG PET has been shown to predict the outcome of tumours, including oropharyngeal cancers[185]. The product of MTV and SUV_{mean} of ^{18}F-FDG PET images is termed the Total Lesion Glycolysis (TLG). This methodology has been applied to ^{18}F-Fluoride imaging where Total Lesion Activity (TLA) = MTV x $SUV_{mean.}$

$SUV_{entropy}$ and SUV_{energy} (also known as uniformity) as heterogeneity parameters were recommended by Hatt et al. following examination of PET textural features for robustness of measurement with respect to the VOI definition method, and limitation of the effect of partial volume impact[186].

The following PET quantification parameters have been selected for these studies:

- **SUV Histogram Parameters**
    - **SUV_{mean}** -the average of voxel SUV values within the VOI.
    - **SUV_{max}** – the value of the highest value SUV voxel within the VOI.
    - **SUV_{peak}**– the maximum average SUV within the hottest 1cm^3 sphere within the VOI

- **SUV Volumetric Parameters**
  - **MTV** – Metabolic Tumour Volume is the volume (cm^3) of the VOI.
  - **TLA** – = MTV x $SUV_{mean.}$
- **SUV Heterogeneity Parameters**
  - **$SUV_{entropy}$** - $SUV_{entropy}$ is a measure of the randomness of the distribution. A high entropy suggests the distribution is spread more widely; a simple image has a low entropy.

$$SUV_{entropy} = -\sum_{b=1}^{L} P(b)\log_2[P(b)]$$

  - **SUV_{energy}** – SUV_{energy} is a 1st order histogram heterogeneity measurement of the uniformity of the distribution. A high value suggests the distribution is more uniform.

$$SUV_{energy} = \sum_{b=1}^{L} [P(b)]^2$$

L=number of bins (=32)

P(b)=probability distribution of bin b

- **SUV Kinetic Parameter**
  - **K_i** – As previously described

2.2.3.2 DW-MRI Quantification Parameters for Analysis

Most centres report the ADC_{mean} and ADC_{median} as standard parameters, and both these have been reported in this analysis. Both describe the VOI ADC histogram, but are affected differently by the distribution shape.

ADC_{min} was not included in the analysis; initial data showed ADC_{min}=0 for most lesions, most likely because of the impact of image noise. A similar finding has been reported by other groups.

The total diffusion volume (tDV) is a parameter reported by other groups[187] and is analogous with the MTV obtained from PET imaging. This is the volume of the VOI(s) as defined on extrapolated b-value image.

$ADC_{entropy}$ and ADC_{energy} heterogeneity parameters have been associated with prostate cancer aggressiveness [176,177] and with prediction of treatment response of primary breast tumours[172] and have been selected for analysis in this study.

ADC Quantification Features – Definitions

- **ADC Parameters**
  - ADC_{mean} - the average of voxel ADC values within the VOI.
  - ADC_{median} - the median of voxel ADC values within the VOI.
- **tDV** – total diffusion volume = volume of VOI.

- **ADC Heterogeneity Parameters**
    - $ADC_{entropy}$ - a measure of the randomness of the distribution. A high entropy suggests the distribution is spread more widely; a simple image has a low entropy.

$$ADC_{entropy} = -\sum_{b=1}^{L} P(b)\log_2[P(b)]$$

    - ADC_{energy} –a 1st order histogram heterogeneity measurement of the uniformity of the distribution. A high value suggests the distribution is more uniform.

$$SUV_{energy} = \sum_{b=1}^{L} [P(b)]^2$$

L=number of bins (=32)

P(b)=probability distribution of bin b

2.2.3.3 Quantifying Change in Parameter

The percentage change (%Δ) in the parameters between scans has been calculated.

$$\%\Delta\ Parameter = 100 \times \frac{(Parameter\ Scan\ 2 - Parameter\ Scan\ 1)}{Parameter\ Scan\ 1}$$

A %Δ of 50% represents a parameter measured at the second scan 50% larger than the baseline scan (baseline scan x 150%) and a second scan parameter smaller than the baseline scan will therefore be represented by a negative change.

2.2.4 Clinical response definitions

2.2.4.1 Prostate Cancer

The clinical response to treatment will be made at up to 24 weeks. Defining PD will be guided by the PCWG2 criteria (Tbl.2-2). Assessment of response will be made by the patient's primary oncologist (blinded to the study results), and a consensus confirmed by a second oncologist.

Feature	Definition of Progression
PSA	*If PSA decline from baseline* • ≥25% and ≥2ng/mL above nadir, confirmed at least 3 weeks later *If no PSA decline from baseline* • ≥25% and ≥2ng/mL after 12 weeks (NB - Ignore early rises in PSA before 12 weeks of treatment, unless other evidence of progression)
Soft Tissue Lesions	As RECIST criteria
Bone Lesions	New lesions confirmed on repeat scan at least 6 weeks later
Symptoms	Clinical suspicion confirmed by continuation/progression of symptoms with second review at least 3 weeks later

Table 2-3: Summary of PCGW2 Guidelines for defining progressive disease[132]

2.2.4.2 Breast Cancer

Radiological assessment will follow MDA criteria[34,148] (using bone scintigraphy, CT, MRI or X-ray as clinically relevant undertaken at 24 weeks, unless clinically indicated before, in consensus by 2 radiologists blinded to PET, DW-MRI and clinical response). The standard imaging at Guys & St Thomas' Hospitals is by bone scintigraphy complemented with X-rays or CT bone windows at 6 months. MRI is used additionally in some patients for monitoring vertebral disease but this combination is applicable to the MDA assessment criteria.

2.2.5 Definitions for Response Analysis

Response (non-progression) – Stable disease or PR/CR (by tumour-specific criteria discussed above) sustained at 24 weeks after initiation of therapy. Stable disease has been included in this response group as clinically none-progression of disease is relevant and will be a clinical reason to continue a chosen therapy.

Progression - Clinically confirmed PD by 24 weeks after starting therapy, determined by the imaging criteria above, or by clinical assessment (tumour markers, clinical signs, clinical assessment) resulting in cessation of ineffective therapy, or a change of treatment.

2.2.6 Definitions for Survival Analyses

Overall survival (OS) is measured as the time, in days, between the date of the first baseline scan for the patient until the data of death, or the date of data censor at the last recorded patient clinical contact.

Progression free survival (PFS) is measured as the time, in days, between the date of the first baseline scan for the patient until the data of clinically confirmed disease progression by the treating clinical team, or the date of data censor at the last recorded patient clinical contact.

2.2.7 General Statistical Analysis

All statistical analysis has been completed in IBM SPSS version 23. Original data was collated and calculated in Microsoft Excel before transfer into SPSS. SPSS has been used to create the figures.

2.3 Methods Specific to Results Chapters

To meet the aims of this analysis, the analysis has been separated into separate sections:

1- ^{18}F-Fluoride PET

   a. ROI Segmentation

   b. Per-Lesion Analysis – Response Analysis

   c. Per-Lesion Analysis – Survival Analysis

   d. Per-Patient Target Lesion Analysis – Impact of VOI segmentation thresholds

   e. Per-Patient Target Lesion Analysis – Response Analysis and impact of VOI segmentation thresholds

   f. Per-Patient Target Lesion Analysis – Survival Analysis

   g. Whole-Body Quantification of Bone Metastases – Methodology, Response Analysis and Survival Analysis

2- DW-MRI

   a. Per-Lesion Analysis – Response Analysis

   b. Per-Lesion Analysis – Survival Analysis

   c. Per-Patient Target Lesion Analysis – Response Analysis

   d. Per-Patient Target Lesion Analysis – Survival Analysis

   e. Whole-Body Quantification of Bone Metastases – Methodology, Response Analysis and Survival Analysis

These specific methods will now be described separately.

2.3.1 ^{18}F-Fluoride PET Segmentation –SUV threshold for defining the PET ROI of bone metastases

Sclerotic bone metastases can be identified on CT imaging as areas on increased density within bone, with clear and often sharp contrast with the surrounding normal bone, and the sclerosis can be used as a proxy for tumour cells within the bone. Taking this assumption, I have worked to identify the best relative threshold to apply to an ^{18}F-Fluoride PET ROI containing a bone metastasis ensuring the volume closely matches the sclerotic region on CT imaging.

Patients with bony metastases from prostate cancer had ^{18}F-Fluoride PET scans as part of the FAP-P and FAB-IE studies previously described. The first four of these scans have been used.

OsiriX (Pixmeo) software was used for ROI definition, for applying the threshold, and for calculating the ROI.

The reference standard is the area of sclerotic change of bone metastasis as identified on the non-contrast enhanced CT component of the ^{18}F-Fluoride PET/CT scan. The largest lesions, clearly distinguished from adjacent lesions and from normal surrounding bone were selected, ensuring there was also tracer avidity on the fused PET component. Up to 5 lesions per patient were delineated on the CT scans, and a record made of the area of the ROI^{CT}.

Manual ROI^{PET} regions were then drawn around the corresponding tracer avid lesions on the PET scan, ensuring a rim of non-avid tissue was included in the ROI^{PET}, and ensuring correct fusion with the CT.

Thresholds relative to the SUV_{max} were then applied to the ROI^{PET}s. The threshold was increased in stages, making note of the resulting ROI^{PET} area. This data was plotted graphically, comparing the volume of the ROI^{PET} (expressed as % of ROI^{CT} volume) against the threshold applied. The line of best fit was applied using Microsoft Excel functionality, and the equation of this best fit line was exported. For each lesion the threshold level when ROI^{PET} matches the corresponding ROI^{CT} was identified. The population of lesions can then be interrogated to identify the best threshold.

Figures 3-1/3-2/3-3/3-4 document all the results, showing the decrease in area of the ROI^{PET}s as the relative threshold applied to the ROI is increased. For many of the lesions this is not a linear relationship.

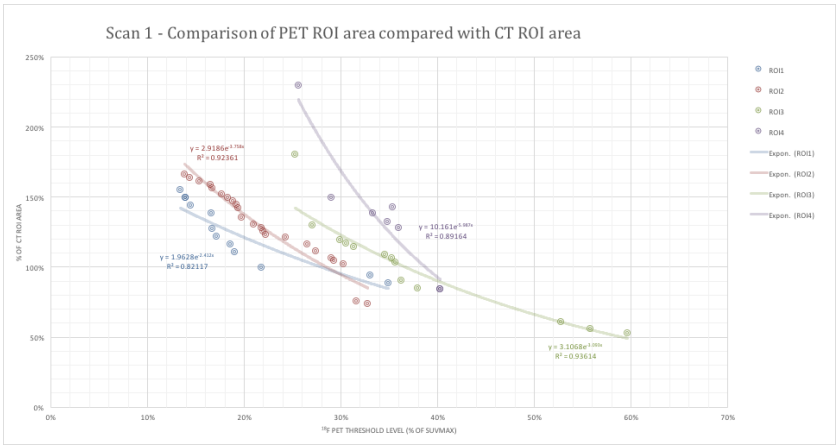


Figure 2-1: Patient 1. Comparison of ROI^{PET} to ROI^{CT} (%) against the relative threshold (%SUV$_{max}$) used to delineate ROI^{PET}

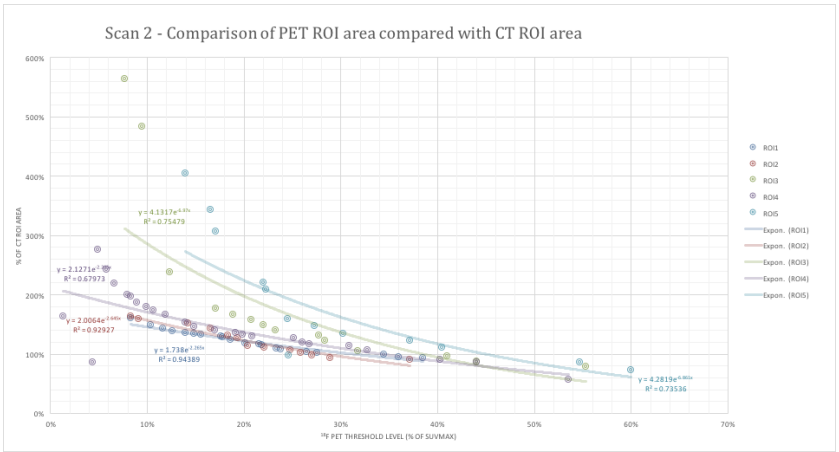


Figure 2-2: Patient 2. Comparison of ROI^{PET} to ROI^{CT} (%) against the relative threshold (%SUV$_{max}$) used to delineate ROI^{PET}

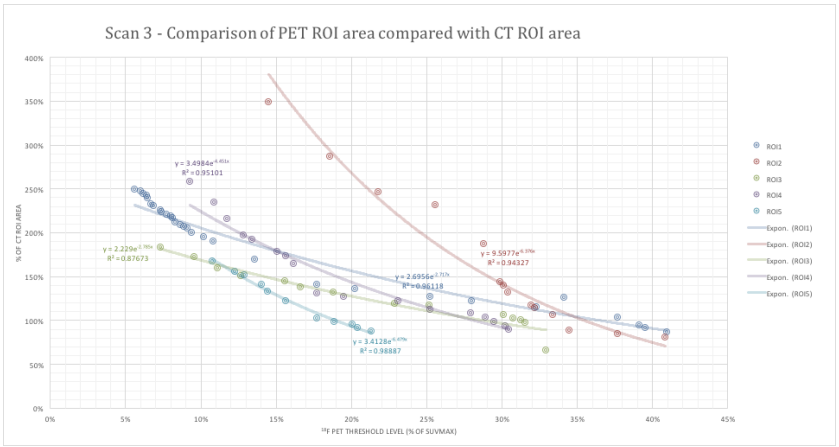


Figure 2-3: Patient 3. Comparison of ROI^{PET} to ROI^{CT} (%) against the relative threshold (%SUV$_{max}$) used to delineate ROI^{PET}

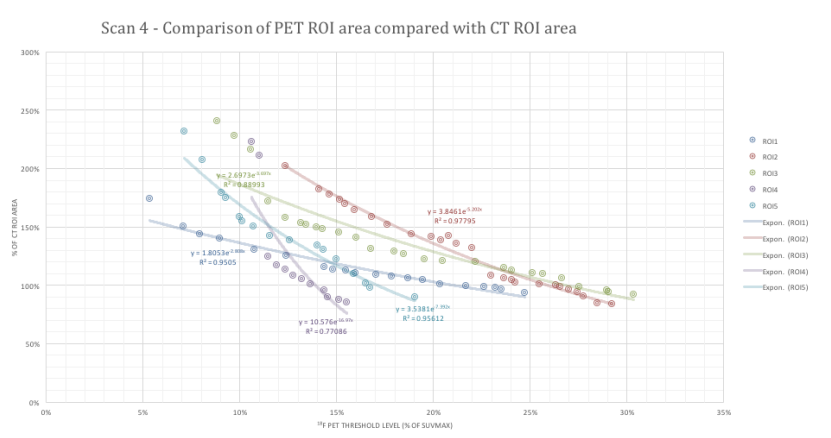


Figure 2-4: Patient 4. Comparison of ROI^{PET} to ROI^{CT} (%) against the relative threshold (%SUV$_{max}$) used to delineate ROI^{PET}

From the lines of best fit it is possible to calculate the relative threshold needed to ensure ROI^{PET} equals ROI^{CT} for each lesion. These data are displayed in the table below. In this sample the mean threshold of ROI^{PET} to match the area of the ROI^{CT} is 27%.

ROI	ROI^{CT} Area (cm^2)	PET threshold of to match ROI^{PET} to ROI^{CT} (%SUV^{max})
P1_ROI1	5.393	27.96%
P1_ROI2	12.618	28.50%
P1_ROI3	11.268	36.65%
P1_ROI4	13.804	38.73%
P2_ROI1	19.414	24.40%
P2_ROI2	7.309	26.33%
P2_ROI3	3.398	20.35%
P2_ROI4	8.981	34.23%
P2_ROI5	2.438	21.20%
P3_ROI1	12.721	36.22%
P3_ROI2	8.139	35.47%
P3_ROI3	14.063	28.78%
P3_ROI4	6.378	28.14%
P3_ROI5	7.866	18.95%
P4_ROI1	17.232	21.04%
P4_ROI2	6.351	25.90%
P4_ROI3	6.491	26.66%
P4_ROI4	9.74	13.90%
P4_ROI5	1.837	17.09%
	Mean	**26.87%**
	SD	**7.09%**
	95% CI	**12.70%-41.04%**

Table 2-4: PET threshold calculated to ensure ROI^{PET}=ROI^{CT}

Taking an average of the ideal threshold across these 19 lesions analysed suggested the segmentation of the manual PET^{ROI} by 27% of SUV$_{max}$ will, on average, equate best to the 'gold-standard' of the CT^{ROI}.

2.3.2 ^{18}F-Fluoride PET – Per-Lesion Analysis – Response Analysis

Hypothesis – Quantification parameters will be shown to predict 24-week clinical treatment response both from the baseline imaging and from the change between scans.

VOI definitions

The largest individually definable bone metastases were selected for VOI definition, ensuring the lesions were also definable on DW-MRI sequences. These same lesions were then identified on the second scan, using the baseline scan and CT component of the PET/CT scan to ensure anatomical accuracy.

Manual VOIs were drawn around each lesion on the PET scan using the registered CT image to ensure anatomic accuracy. In-house software (FAST) was used to define the volumes. The manual VOI (VOI_{manual}) was drawn by the same investigator (Dr Benjamin Taylor (author of this thesis) Clinical Oncologist with practical radiology experience) to ensure full inclusion of the PET avid area within the VOI, and ensuring physiological tracer uptake was excluded (confirmation with an experienced Nuclear Medicine physician where necessary).

An iso-contour of 27% of SUV_{max} was applied, excluding peripheral voxels of less than 27% (VOI_{27}) (see Chapter 2.3.1).

The same lesion on the second scan was identified using all available clinical imaging modalities and the CT component of the PET scan to ensure anatomical accuracy of delineation.

PET Quantification Features - Definitions

- **SUV Histogram Parameters –** SUV_{mean}, SUV_{max}, SUV_{peak}
- **SUV Volumetric Parameters -** MTV, TLA.
- **SUV Heterogeneity Parameters -** $SUV_{entropy}$, SUV_{energy}.
- **SUV Kinetic Parameter** - K_i.

The parameters are reported from the baseline scan, the second scan (at 8 weeks), and %Δ.

Statistical Analyses

- Comparison of parameters between tumour sub-groups using parametric independent samples t-tests (Mann-Whitney U tests if assumptions necessary for a parametric approach are not met)
- Comparison of parameters between the 24-week response groups using independent samples t-tests (Mann-Whitney U tests if assumptions necessary for a parametric approach are not met)

Results - See Chapter 3

2.3.3 ^{18}F-Fluoride PET – Per-Lesion Analysis – Survival Analysis

Hypothesis – Quantification parameters from the baseline scan and changes in parameters between scans will be identified to be associated with OS and PFS.

VOI definitions - as described in “Per-Lesion Analysis – Response Analysis” section

PET Quantification Features- as described in “Per-Lesion Analysis – Response Analysis” section

Statistical Analyses

- OS Analysis
    - Cox regression univariate analysis
    - KM Analysis with log rank analysis
- PFS Analysis
    - Cox regression univariate analysis
    - KM Analysis with log rank analysis

Results – See Chapter 4

2.3.4 ^{18}F-Fluoride Per-Patient Target Lesion Analysis – Impact of VOI segmentation thresholds

Hypothesis – Choice of VOI will have a significant impact on the quantification parameters, and it may be possible to identify a method with the lowest variation. %Δ parameters will less variation with VOI method.

VOI definitions

Examining the impact of VOI segmentation on quantification is a major component of this analysis, and therefore a number VOI definition methods have been developed.

The largest individually definable bone metastases were selected for VOI definition, ensuring the lesions were also definable on DW-MRI sequences. Up to 5 lesions were identified for each patient. These same 5 lesions were then identified on the second scan.

Manual VOIs were drawn around each lesion on the PET scan using the registered CT image to ensure anatomic accuracy. In-house software (FAST) was used to define the volumes. Further VOIs were generated based on the SUV_{max} of each VOI_{manual}, excluding voxels at less than 27% (VOI_{27}) (based on preliminary study described elsewhere in this thesis), 40% (VOI_{40})[100,101] and 50% (VOI_{50})[188] of the VOI SUV_{max} to create iso-contour SUV threshold VOIs for the baseline imaging. A VOI using a FLAB technique was also created (VOI_{FLAB})[113]. When a volume was fragmented into separate volumes by application of an SUV segmentation threshold, the largest residual volume was considered as the VOI for quantification to minimise the impact of image noise on quantification (greater impact on smaller VOIs).

A further set of absolute SUV threshold VOIs were created on the second PET scan using the absolute SUV thresholds used to create the baseline (BL) scan VOIs (termed VOI_{27BL}, VOI_{40BL}, $VOI_{50BL,}$ respectively). There is no consensus in published literature about how SUV parameters based on the whole voxel population of a VOI, or volumetric parameters, might be used to assess changes overtime. Tracer uptake in a VOI is heterogeneous, and therefore using separate iso-contours on first and second scans may provide misleading SUV volumetric measurements. An absolute contour based on the iso-contour SUV of the baseline scan has therefore been investigated.

e.g. Baseline Scan – VOI right femur. VOI_{manual} has 40% of SUV_{max} threshold applied to create VOI_{40}. For this example, 40% of SUV_{max} = 4.

Second Scan – same VOI right femur identified and VOI_{manual} drawn. An absolute SUV threshold of 4 used to create $VOI_{40BL.}$ In addition, a VOI_{40} is also generated using the iso-contour threshold method.

Baseline Scan	Second Scan	Type of segmentation method
VOI_{manual}	VOI_{manual}	Manual
VOI_{27}	VOI_{27}	Iso-contour
VOI_{40}	VOI_{40}	
VOI_{50}	VOI_{50}	
VOI_{FLAB}	VOI_{FLAB}	Adaptive
	VOI_{27BL}	Absolute
	VOI_{40BL}	
	VOI_{50BL}	

Table 2-5: Summary of VOI methods used for quantification PET Quantification Features

When more than one VOI has been identified in a patient the average of the lesions (up to 5) has been used for quantification, with the exception of MTV where the sum of up to 5 lesions has been calculated.

- **SUV Histogram Parameters –** SUV_{mean}, SUV_{max}, SUV_{peak}
- **SUV Volumetric Parameters -** MTV, TLA.
- **SUV Heterogeneity Parameters -** $SUV_{entropy}$, SUV_{energy}.
- **SUV Kinetic Parameter** - K_i.

The parameters are reported from the baseline scan, the second scan (at 8 weeks), and %Δ.

Statistical Analyses

- Descriptive statistics
- Comparisons between parameter distributions calculated with each VOI method using a parametric repeated measures one-way ANOVA (where the assumptions necessary for a parametric approach are not met, a non-parametric Friedman's Variance test)
- Comparisons between the tumour groups with Mann-Whitney U tests

Results – See Chapter 5

2.3.5 ^{18}F-Fluoride Per-Patient Target Lesion Analysis – Response Analysis and Impact of VOI Segmentation Thresholds

Hypothesis – Quantification parameters from the baseline scan and changes in parameters between scans will be identified predict 24-week clinical treatment response.

VOI definitions

The VOI were defined as outlined in the previous sections.

PET Quantification Features – Definitions - as described in "^{18}F-Fluoride Per-Patient Target Lesion Analysis – Impact of VOI segmentation thresholds" section

Statistical Analyses

- Mann-Whitney U tests to compare distributions of parameters between response groups for each VOI method separately.

Results – See Chapter 6

2.3.6 ^{18}F-Fluoride Per-Patient Target Lesion Analysis – Survival Analysis

Hypothesis – Quantification parameters from the baseline scan and changes in parameters between scans will be identified to be associated with OS and PFS.

VOI definitions – as described in "^{18}F-Fluoride Per-Patient Target Lesion Analysis – Impact of VOI segmentation thresholds" section. It was not reasonable to perform survival analyses using all VOI methods. This analysis has been carried out on the VOI^{manual}.

PET Quantification Features – Definitions - as described in "^{18}F-Fluoride Per-Patient Target Lesion Analysis – Impact of VOI segmentation thresholds" section

Statistical Analyses

- OS Analysis
    - Cox regression univariate analysis
    - KM Analysis with Log Rank Analysis
- PFS Analysis
    - Cox regression univariate analysis
    - KM Analysis with Log Rank Analysis

Results – See Chapter 7

2.3.7 ^{18}F-Fluoride Whole-Body Quantification of Bone Metastases – Methodology, Response Analysis and Survival Analysis

Hypothesis – Parameters representing the whole PET skeletal volume will be quantified. Parameters will be identified predicting 24-week treatment response, and prognostic of PFS and OS.

VOI definitions

A method was developed for this analysis to segment the bony skeleton from the CT component of the PET/CT scan, and use this as a mask to identify and isolate the PET data only corresponding to the skeleton. The scans were manually edited to exclude any residual non-skeletal or physiological PET signal.

Software for auto-segmentation of the skeleton from the non-skeletal CT data was not available, and therefore this methodology was developed using readily available free software (OsiriX, for Mac). Bone is readily distinguished on CT images due to the significant difference in density of bone compared to the surrounding tissues.

The baseline CT components of the PET/CT scans of 3 breast cancer patients and 4 prostate cancer patients (the first patients recruited to the studies already described) were used for analysis. Multiple ROIs were drawn within skeletal and non-skeletal tissues, and the Hounsfield Unit distribution statistics for each ROI recorded.

	Skeletal ROI	Extra-skeletal ROI
Mean Hounsfield Unit (HU) (95% CI)	288.88 (262.84-314.92)	-4.891 (-16.496-6.714)
HU Standard Deviation (95% CI)	167.614 (152.988-182.241)	28.804 (26.572-31.035)
HU Min (95% CI)	-56.532 (-78.84 - -34.225)	-92.899 (-109.905 - -75.89)
HU Max (95% CI)	862.771 (816.040-909.502)	75.827 (63.948-87.707) SD=70.834

Table 2-6: Population statistics of Hounsfield Unit (HU) descriptors for skeletal and extra-skeletal ROIs.

To differentiate between skeletal and extra-skeletal tissues a HU threshold greater than the maximum HU of extra-skeletal tissue, and ideally less than the minimum HU of skeletal bone needs to be identified.

For this methodology, the threshold selected was the maximum HU of the extra-skeletal ROI population, plus 2 standard deviations (where SD of population =70.834).

Therefore, the threshold identified was 75.827+2(70.834)=217.5.

This threshold will exclude 95% of extra-skeletal uptake. Visual inspection identified a good segmentation of the skeletal volume.

Skeletal Segmentation Method

The following details the step-by-step CT processing used to segment the skeletal volume, creating a mask to be applied to the ^{18}F-Fluoride PET for whole-body skeletal quantification. OsiriX was used to prepare the CT scan, removing any non-skeletal component of the CT scan with HU>217.5. In-house software (FAST) was used to create the mask. The following table summarises the steps involved:

Step	Task
1	CT component of PET scan loaded into OsiriX
2	All pixels with HU<217.5 set to -1000
3	Axial CT views – Residual non-skeletal tissue manually removed (teeth, couch, urinary tract, bladder, injection site, extra-skeletal physiological signal). ROIs manually drawn around all the residual non-skeletal signal, and all these pixels manually set to HU -1000. ROI data saved.
4	Re-check axial views, further deletions if necessary. ROIs of further deletions saved.
5	CT scan closed **without** saving the deletions.
6	Clean CT scan re-opened
7	The saved ROIs were reloaded and the pixels within the ROIs set to -1000. This resulted in all the CT image data being saved, apart from the extra-skeletal tissue which would not be segmented by the 217.5HU threshold. OsiriX applies a universal threshold algorithm whereby voxels below the threshold but within the VOI are also excluded; the FAST software uses an algorithm to apply the segmentation to the edges of the VOI, preserving intra-VOI voxels for quantification.
8	The modified CT was then exported in DIACOM format.
9	FAST software then used. PET component and CT scan loaded into software.
10	Mask created from the modified CT scan, but excluding all pixels<217.5 HU. This mask was then applied over the fused PET scan, and PET voxels outside of the mask excluded from analysis.
11	Skeletal malignancy was then segmented from the whole skeletal PET volume using the 27% iso-contour thresholds based on the SUV_{max} of the skeletal PET volume

Table 2-7: Preparation of CT-scan for segmentation of whole-skeletal volume which can then be used as a mask to select only skeletal ^{18}F-Fluoride PET activity.

Figure 2-5 (below) documents one of the study patients (the same patient has also been pictorially represented in the discussion of the DW-MRI whole-body methodology), showing the baseline bone scan for reference (a), the bony skeleton segmented from the CT data (b), the baseline ^{18}F-Fluoride PET (c), and the segmented ^{18}F-Fluoride PET after application of the mask and a 27% iso-contour for disease definition (d).

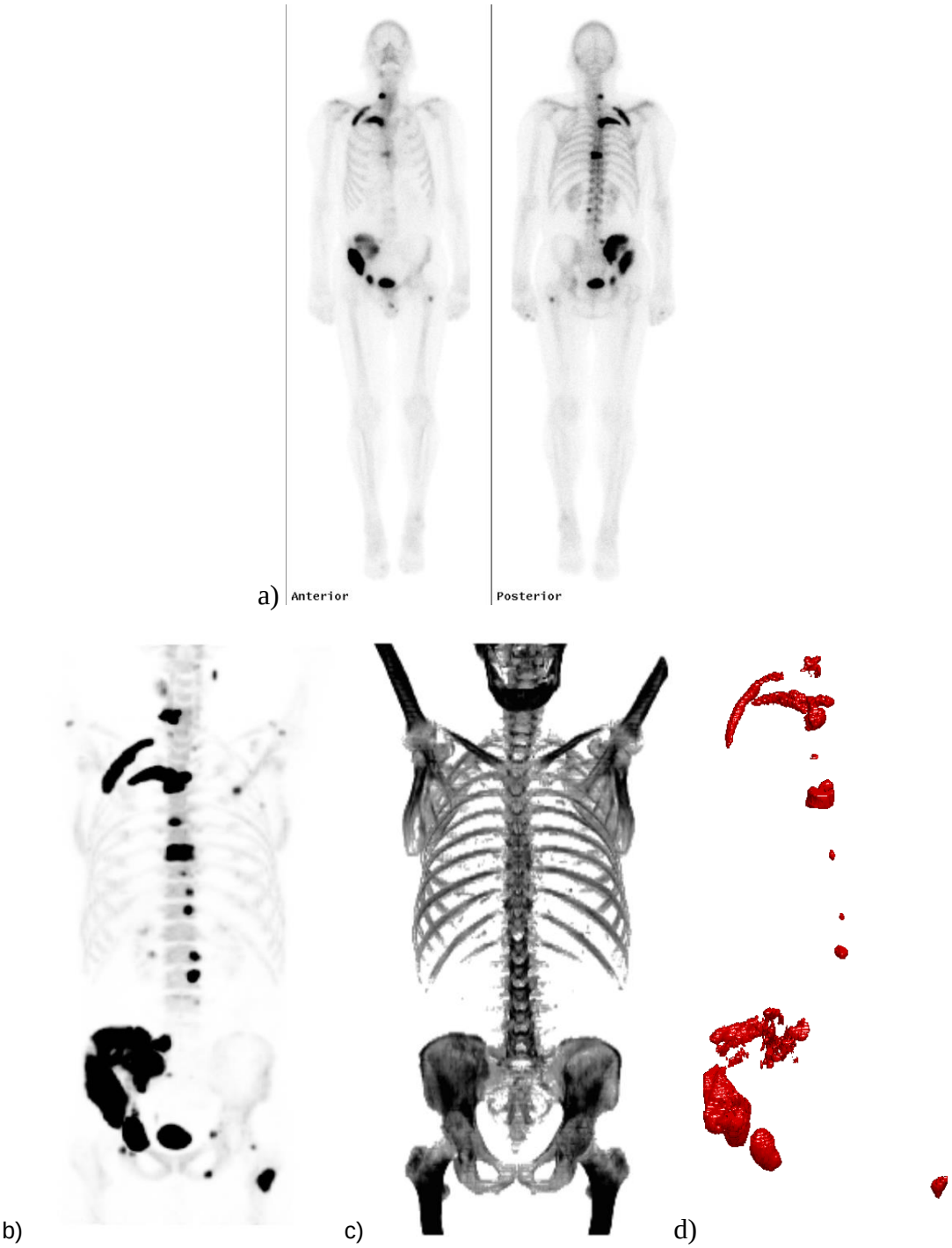


Figure 2-5: Prostate cancer patient demonstrating whole-body segmentation of skeletal disease (see text above for descriptors)

PET Quantification Features – Definitions

For all the patients, there were many VOIs created which had individual quantification parameters. Single pixels were also segmented by the image processing, which may represent clinically relevant bone metastases, or may represent noise. Smaller lesions are more likely to be influenced by partial volume effects and noise. A correction factor was applied to calculate the whole-body parameters (apart from the volumetric parameters) based on the VOI size, with the aim of including all voxels, but reducing the impact of noise on the measured parameter:

$$Parameter_{WB} = \sum VOI\ Parameter\ x\ \frac{VOI\ Volume\ (tDV)}{Volume\ all\ VOIs\ (tDV_{WB})}$$

- **SUV Histogram Parameters –** SUV_{mean}^{WB}, SUV_{max}^{WB}, SUV_{peak}^{WB} (weighted as above)
- **SUV Volumetric Parameters –** MTV_{WB} (sum of all VOIs), TLA_{WB} (MTV^{WB} x SUV_{mean}^{WB}).
- **SUV Heterogeneity Parameters -** $SUV_{entropy}^{WB}$, SUV_{energy}^{WB} (weighted as above)
- **SUV Kinetic Parameter** - K_i^{WB}.(weighted as above)

The parameters are reported from the baseline scan, the second scan (at 8 weeks), and %Δ.

Statistical Analyses

- Descriptive statistics
- Response analysis, comparing distributions between response groups with non-parametric Mann-Whitney U tests (none of data met assumptions necessary for a parametric approach).
- OS and PFS analysis – Cox univariate regression analysis and KM analysis

Results – See Chapter 8

2.3.8 DW-MRI – Per-Lesion Analysis – Response Analysis

Hypothesis – Quantification parameters from the baseline scan and changes in parameters between scans will be identified to be predictive of 24-week treatment response.

VOI definitions

It was felt necessary to ensure the same target lesions were used between imaging modalities. The same largest definable bone metastasis was identified on all imaging modalities, and the MRI sequences used to identify the lesion on DW-MRI imaging, ensuring cross reference between the imaging modalities to ensure anatomical accuracy.

The target lesions were manually delineated using the calculated high-b-value image (b-1400 mm^2/s). No threshold was applied to these volumes before quantification

ADC Quantification Features – Definitions

- **ADC Histogram Parameters –** ADC_{mean}, ADC_{median}
- **ADC Volumetric Parameter -** tDV.
- **ADC Heterogeneity Parameters -** $ADC_{entropy}$, ADC_{energy}.

The parameters are reported from the baseline scan, the second scan (at 8-12 weeks), and %Δ.

Statistical Analyses

- Descriptive statistics of parameters
- Comparison of parameter distributions between response groups with Mann-Whitney U tests

Results – See Chapter 9

2.3.9 DW-MRI – Per-Lesion Analysis – Survival Analysis

Hypothesis – Quantification parameters from the baseline scan and changes in parameters between scans will be identified to be associated with OS and PFS.

VOI definitions

Up to 5 lesions were identified in each patient. The same lesions were identified and quantified on the other imaging modalities, confirming the malignant status of the target lesions. The largest and/or most avid discrete bone lesions were selected for all modalities, ensuring cross reference between the imaging modalities to ensure anatomical accuracy.

The target lesions were manually delineated using the calculated high-b-value image (b-1400 mm^2/s). No threshold was applied to these volumes before quantification

ADC Quantification Features – Definitions – as described in "DW-MRI – Per-Lesion Analysis – Response Analysis" section

Statistical Analyses

- OS Analysis
    - Cox regression univariate analysis
    - KM Analysis with Log Rank analysis
- PFS Analysis
    - Cox regression univariate analysis
    - KM Analysis with Log Rank analysis
- **Results** – See Chapter 10

2.3.10 DW-MRI – Per-Patient Analysis – Response Analysis

Hypothesis – Quantification parameters from the baseline scan and changes in parameters between scans will be identified to be predictive of 24-week treatment response.

VOI definitions

Up to 5 lesions were identified in each patient. The same lesions were identified and quantified on the other study imaging modalities. The largest and/or most avid discrete bone lesions were selected for all modalities, ensuring cross reference between the imaging modalities to ensure anatomical accuracy.

The target lesions were manually delineated using the b=1400 mm^2/s image.

ADC Quantification Features – Definitions

ADC_{mean} - the average SUV_{mean} of VOI(s).

ADC_{median} - the average SUV_{median} of VOI(s).

tDV – total diffusion volume = sum of individual VOI volumes

$ADC_{entropy}$ – the average $SUV_{entropy}$ of VOI(s).

ADC_{energy} – the average SUV_{energy} of VOI(s).

The parameters are reported from the baseline scan, the second scan (at 8-12 weeks), and %Δ.

Statistical Analysis

- Response Assessment
  - Non-parametric Mann-Whitney U test to compare the parameter distributions between response groups (a non-parametric approach is used because of small sample size)

Where appropriate, generation of Receiver Operating Characteristic (ROC) curves.

Results See Chapter 11.

2.3.11 DW-MRI – Per-Patient Analysis – Survival Analysis

Hypothesis – Quantification parameters from the baseline scan and changes in parameters between scans will be identified to be associated with OS and PFS.

VOI definitions – see Chapter 2.3.10

Statistical Analysis

- OS Analysis
    - Cox regression univariate analysis
    - KM Analysis with Log Rank test for comparison
- PFS Analysis
    - Cox regression univariate analysis
    - KM Analysis with Log Rank test for comparison

Results – See Chapter 12

2.3.12 DW-MRI Whole-Body Quantification of Bone Metastases – Methodology, Response Analysis and Survival Analysis

Hypothesis – Parameters representing the whole PET skeletal volume will be quantified. Parameters will be identified predicting 24-week treatment response, and prognostic of PFS and OS.

It has been necessary to develop a method for identifying and defining the full skeletal metastatic burden to enable whole-body quantification of the bone metastases.

Extrapolated b=1400s/mm^2 images were generated for each DW-MRI scan. This high diffusion gradient strength exaggerates the signal from restricted diffusion by suppressing the T2 shine through effect, reducing the influence of capillary perfusion on the image signal[189]. A high b-value DW-MRI image produces stronger diffusion-weighting and more suppression of benign tissue signal. have proved their utility for detecting malignancy. These images were therefore selected for VOI^{WB} definition.

It is not currently feasible to segment the skeleton out of an MRI image, and therefore I have worked to find a solution for whole-body selection of only bony metastatic disease. This is complicated by the significant physiological signal on DW-MRI scanning, and still present on the high b-value images.

A small preliminary study involved the manual delineation of 14 known bone metastases. These were confirmed as sites of metastatic disease in prostate cancer patients, and CT imaging was used to identify the area of sclerosis. Bone lesions clearly distinguishable from surrounding normal bone and surrounding structures were identified. The corresponding anatomical sites were identified on the DW-MRI imaging, using all sequences to aid identification of the site of disease. A manual VOI was then drawn around the lesion, using Osirix on an Apple Mac computer. The volumes of the CT defined sclerotic disease was compared with the manual high-b signal VOI with a paired samples t-test, showing where t(8)=-1.338 p=0.218, concluding there was no significant difference in the volume of disease characterised on each imaging modality. The high-b value image signal data was analysed by comparing the signal intensity maximum. The data can be seen in the following table and box-plot comparing the benign and malignant lesions.

Inclusion of all malignant tissue was identified as the priority, rather than exclusion of all benign tissue. The distributions of the high b-value image overlap, and therefore there is no perfect distinguishing cut-off. Selecting a high b-value imaging intensity of 14 was identified on ROC analysis as the most differentiating value, with a sensitivity of 100% (meaning all malignant lesions would be identified if the VOI contained a high b-value signal intensity of 14).

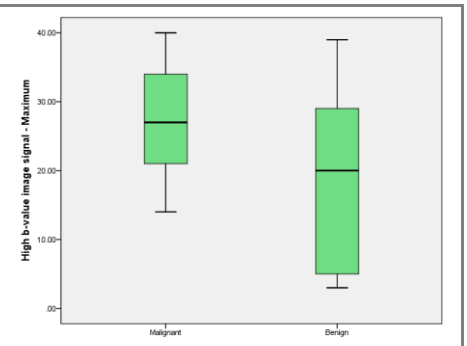
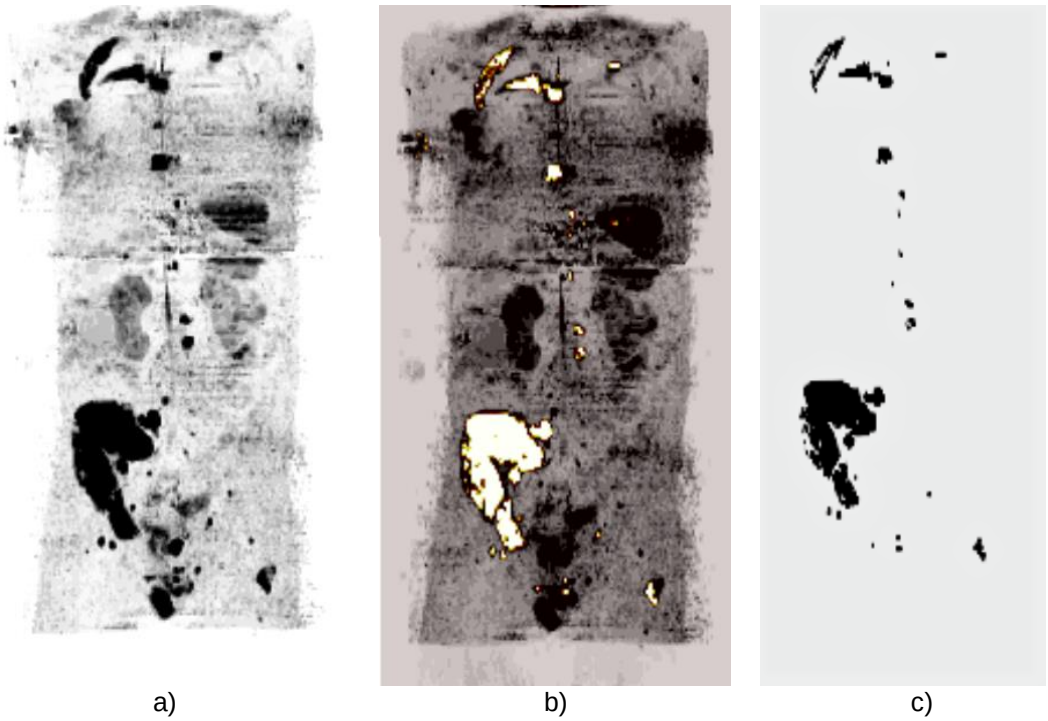
	High b-value image signal	
	Mean (min-max) [median; SD]	
All lesions n=24	**21.91** (3.0-40.0) [22.0; 11.7]	
Malignant n=14	**27.11** (14.0-40.0) [27.0; 8.55]	
Benign n=14	**18.57** (3.0-39.0) [20.0; 12.49]	

Table 2-8: High b-value DW-MRI imaging analysis; benign vs malignant bone lesions

For this feasibility study this threshold was used for initial image preparation, but the high b-value images required manual deletion of physiological and non-malignant image signal. Firstly, all voxel values in the high b-value image were made binary: all voxels of intensity<14 were set to 0 (black), and all remaining voxels set to 1000 (white) to ensure all residual voxels were clearly distinguishable. Slice by slice non-malignant non-bone signal was deleted, using other sequences image-fused for anatomic guidance.

The final modified b-value image was then used as a mask to place over the ADC map, and each lesion was then quantified using in-house image analysis software.



a) b) c)

Figure 2-6: Whole-body DW-MRI VOI Segmentation: Study Patient FAP03 Baseline Scan. (^{99m}TC Bone Scan showing sites of prostate cancer bone metastases and ^{18}F-Fluoride PET scan showing uptake at sites of bone metastases can be reviewed in whole-body ^{18}F-Fluoride methodology) a) extrapolated high b-value (1400s/mm2) image of same patient b) high-b value image with segmented pathological bone signal mask overlaid c) pathological bone signal mask which will be applied to ADC map for whole-body quantification)

ADC Quantification Features – Definitions

The images after processing demonstrated image noise, including numerous single pixels. Some studies have only quantified lesions >2cm. For this study a weighting was applied to create the Whole-body (WB) parameter, minimising the impact of individual pixels/small volumes more likely to represent image noise.

$$Parameter_{WB} = \sum VOI\ Parameter\ x\ \frac{VOI\ Volume\ (tDV)}{Volume\ all\ VOIs\ (tDV_{WB})}$$

ADC_{mean}^{WB} – As previously described, weighted as above

ADC_{median}^{WB} - As previously described, weighted as above.

tDV_{WB} – total diffusion volume = volume of VOI(s)

$ADC_{entropy}^{WB}$ – As previously described, weighted as above

ADC_{energy}^{WB} – As previously described, weighted as above

The parameters are reported from the baseline scan, the second scan (at 8-12 weeks), and %Δ.

Statistical Analysis

- Tumour group analysis
    - Comparison of distributions of parameters between tumour groups made using independent samples t-tests (where assumptions for a parametric approach were not met, Mann-Whitney U tests used)
- Response Analysis
    - Comparison of distributions of parameters between response groups made using independent samples t-tests (or non-parametric, Mann-Whitney U tests)
- OS Analysis
    - Cox regression univariate analysis
    - KM Analysis
- PFS Analysis
    - Cox regression univariate analysis
    - KM Analysis

Results – See Chapter 13

2.4 Patients

At the time of this data analysis, 20 patients had been recruited to the three studies. The patient demographics and clinical details can be seen in the summary table on the following page.

5 women with breast cancer have been included, all of whom were treated with hormone therapies and bisphosphonates. The average Ca15-3 was 167 (range 14-484). The breast cancer patients were of average age 65 years (range 40-78 years).

15 men with prostate cancer have been included. 6 patients were treated with docetaxel chemotherapy, and 9 with Abiraterone and prednisolone. The average PSA was 308 (range 7-2891). The prostate cancer patients were of average age 74 years (range 60-90 years).

14 patients had baseline ^{18}F-Fluoride PET imaging, and 12 successfully completed the second imaging, with an average 69 days between the scans (range 53-93 days).

20 patients had baseline DW-MRI imaging, and 19 successfully completed the second imaging, with an average of 71 days between the scans (range 55-86 days).

Patients

PT	Study	Tumour	Age (yrs)	Treatment	Days between 1st scan and starting treatment	Baseline Ca153	Baseline PSA	Scan1 ^{18}F-Fluoride	Scan2 N^{18}F-Fluoride	Days Between ^{18}F-Fluoride	MRI1	MRI2	Days Between MRI	Dead/Alive	Progressed/Not Progressed	24Week Response
1	FAB-B	Breast	40	Letrozole and bisphosphonates	13	484	N/A	✓	✓	57	10/07/13	03/09/13	55	Alive	Progressed	No Response
2	FAB-B	Breast	75	Letrozole and zometa	33	14	N/A	✓	✓	58	09/09/13	06/11/13	58	Alive	Not Progressed	Response
3	FAB-B	Breast	68	Letrozole + zometa	7	47	N/A	✓	✓	62	30/10/13	07/01/14	69	Alive	Not Progressed	Response
4	FAB-B	Breast	78	Tamoxifen	14	87	N/A	✓	✓	62	20/01/14	20/03/14	59	Alive	Progressed	No Response
5	FAB-B	Breast	65	Faslodex+Denosumab	15	205	N/A	✓	✓	77	07/04/14	18/06/14	72	Alive	Not Progressed	Response
6	FAB-P	Prostate	66	Docetaxel (+GCSF)	2		475	✓		n/a	01/10/12	03/12/12	63	Dead	Progressed	No Response
7	FAB-P	Prostate	72	Docetaxel	4		2891	✓	✓	63	21/01/13	04/04/13	73	Alive	Progressed	No Response
8	FAB-P	Prostate	79	DOCETAXEL	17		173	✓	✓	93	11/02/13	26/04/13	74	Dead	Progressed	No Response
9	FAB-P	Prostate	74	DOCETAXEL	6		137	✓	✓	68	30/08/13	01/11/13	63	Alive	Not Progressed	Response
10	FAB-P	Prostate	90	ABIRATERONE+PRED	4		76	✓	✓	53	28/01/14	25/03/14	56	Alive	Progressed	No Response
11	FAB-P	Prostate	68	DOCETAXEL	18		39	✓	✓	84	24/03/14	06/06/14	74	Alive	Progressed	No Response
12	FAB-P	Prostate	79	ABIRATERONE+PRED	9		76	✓	✓	85	04/04/14	26/06/14	83	Alive	No Progressed	n/a
13	FAB-P	Prostate	77	DOCETAXEL	1		8	✓	✓	70	08/04/14	05/06/14	58	Dead	Progressed	No Response
14	FAB-P	Prostate	64	ABIRATERONE+PRED	21		263	✓	x	n/a	10/04/14	n/a	N/A	Alive	Not reached	Not reached
15	FAB-IE	Prostate	82	ABIRATERONE+PRED	4		157				08/07/13	30/09/13	84	Dead	Progressed	No Response
16	FAB-IE	Prostate	60	ABIRATERONE+PRED	2		29				12/08/13	06/11/13	86	Alive	Progressed	No Response
17	FAB-IE	Prostate	79	ABIRATERONE+PRED	11		7				19/08/13	12/11/13	85	Alive	Progressed	No Response
18	FAB-IE	Prostate	63	ABIRATERONE+PRED	1		213				30/09/13	20/12/13	81	Dead	Progressed	No Response
19	FAB-IE	Prostate	83	ABIRATERONE+PRED	-2		37				12/11/13	28/01/14	77	Alive	Progressed	No Response
20	FAB-IE	Prostate	74	ABIRATERONE+PRED	1		42				20/11/13	14/02/14	86	Alive	Not Progressed	Response

Table 2-9: Patient demographics

Chapter 3 ^{18}F-Fluoride PET – Per-Lesion Analysis. Image quantification for response assessment of patients with bone metastases from breast or prostate cancer

The methods of this chapter and the statistical approaches for this section can be reviewed in Chapter 2.3.2.

3.1 Results

14 patients had data suitable for inclusion, with 62 bone metastases identified. 5 breast cancer patients have yielded 24 bone metastases, and 38 lesions from 9 prostate cancer patients. Response data was available for 55 lesions (12 patients, 24 breast lesions, 31 prostate lesions). 2/5 (40%) breast cancer patients (representing 10/24 lesions) had PD. 6/7 (86%) prostate cancer patients (29/31 lesions) had PD. With only 2 prostate lesions in patients with a treatment response, response analysis of the prostate cancer lesions only is not feasible.

3.1.1 Baseline Scan

3.1.1.1 SUV Parameters

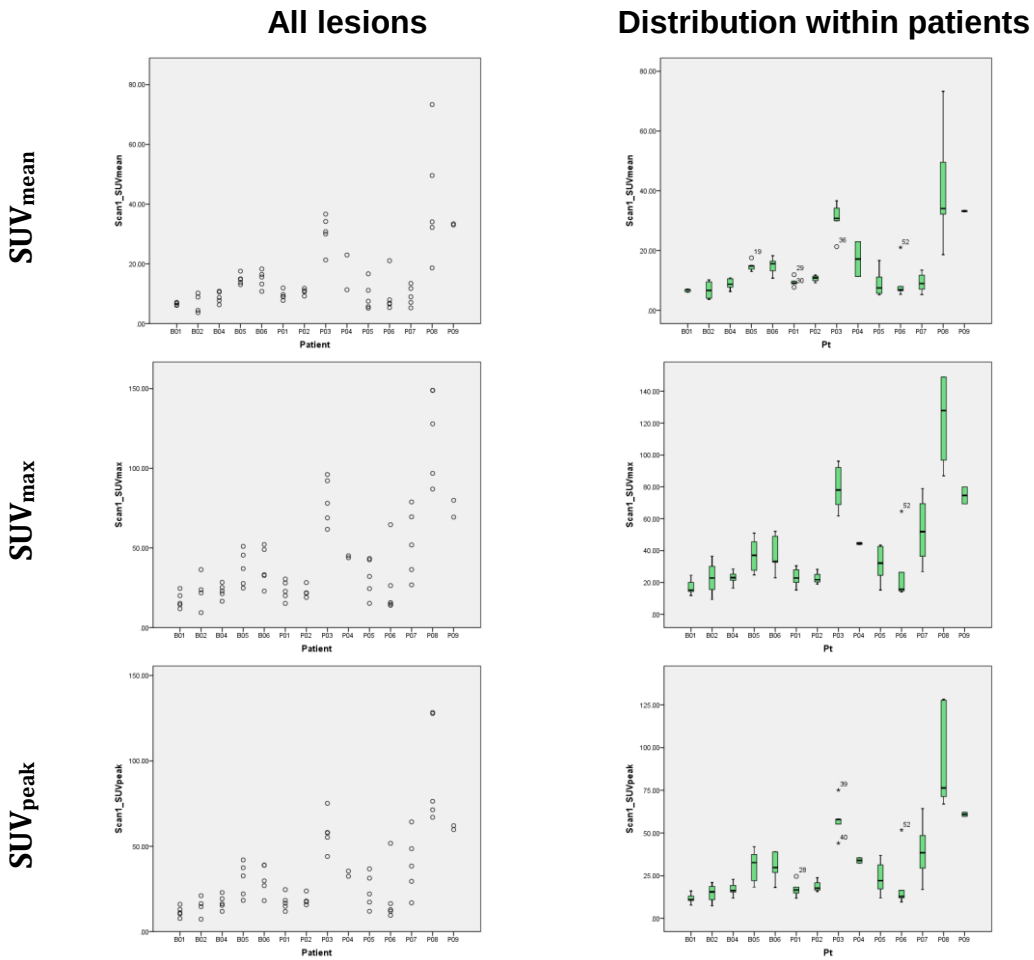


Figure 3-1: ^{18}F-Fluoride Per-lesion analysis; Baseline SUV Parameters

Fig.3-1 demonstrates the heterogeneity of SUV parameters. The whole-population statistics are recorded in Tbl.3-1 and Fig.3-2:

	SUV$_{mean}$	**SUV$_{max}$**	**SUV$_{peak}$**
		Mean (95% CI) [min-max] {median; SD} |Skewness; Kurtosis|	
All lesions n=62	**15.26 (12.13-18.39)** [3.7-73.31] {10.94; 12.33} |2.42; 7.62|	**43.18 (34.99-51.37)** [9.39-148.92] {29.45; 32.25} |1.65; 2.6|	**32.91 (26.39-39.42)** [7.31-128.33] {22.49; 25.66} |1.91; 4.42|
Breast (B) n=24	**10.56 (8.75-12.38)** [3.7-18.29] {10.4; 4.29} |0.24; -1.11|	**27.84 (22.72-32.97)** [9.39-52.19] {24.7; 12.14} |0.69; -0.28|	**21.19 (16.81-25.58)** [7.31-42] {18.24; 10.39} |0.72; -0.58|
Prostate (P) n=38 (7 did not have response analysis data)	**18.22 (13.4-23.05)** [5.2-73.31] {11.56; 14.69} |1.83; 4.15|	**52.87 (40.68-65.05)** [14.01-148.92] {43; 37.07} |1.11; 0.66|	**40.31 (30.6-50.02)** [9.55-128.33] {31.86; 29.55} |1.42; 2.18|
Progression n=39 (B=10 P=29)	**16.74 (12.13-21.34)** [5.2-73.31] {11.15; 14.21} |2.2; 5.84|	**45.62 (33.41-57.83)** [11.86-148.92] {27.99; 37.67} |1.49; 1.45|	**34.75 (25.07-44.42)** [7.73-128.33] {22.08; 29.85} |1.77; 3.14|
Response n=16 (B=14, P=2)	**11.27 (8.55-13.99)** [3.7-22.99] {10.68; 5.1} |0.71; 0.5|	**30.32 (23.81-36.83)** [9.39-52.19] {26.81; 12.22} |0.36; -0.66|	**22.86 (17.64-28.08)** [7.31-38.94] {20.17; 9.8} |0.39; -0.96|

Table 3-1: ^{18}F-Fluoride PET Per-lesion analysis; Descriptive statistics of baseline scan; SUV Parameters

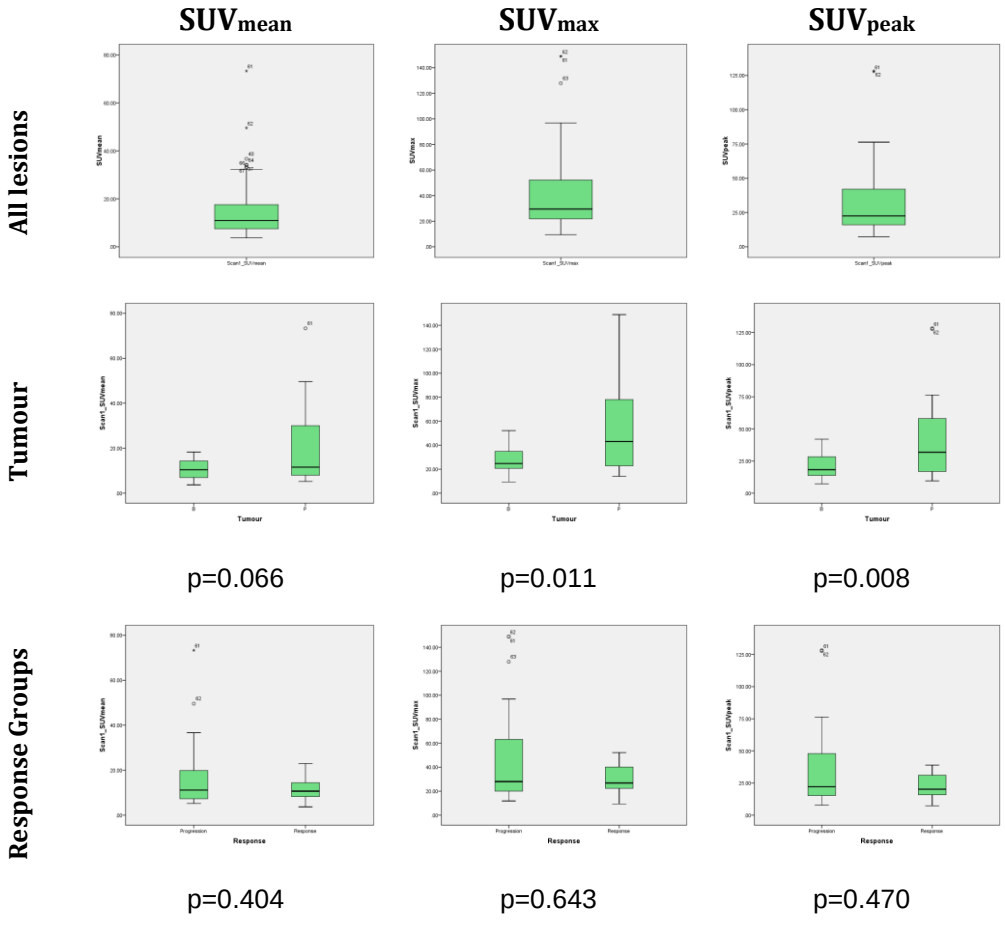


Figure 3-2: ^{18}F-Fluoride PET Per-lesion analysis; Descriptive statistics of baseline scan; SUV Parameters; Comparisons with Mann-Whitney U tests (B - Breast cancer patients; P – Prostate cancer)

The mean of the SUV parameter distributions are significantly higher for the prostate cancer lesions (Mann Whitney U tests: SUV_{mean} p=0.066; SUV_{max} p=0.011; SUV_{peak} p=0.008). Prostate cancer lesions are predominantly osteoblastic, and it is expected they would have increased avidity for ^{18}F-Fluoride. The standard deviation of the groups distributions is smaller for breast cancer lesions (4.29) compared with the prostate cancer lesions (14.69) suggesting more inter-lesion heterogeneity in the prostate cancer lesions measured.

There is no demonstrable association between the baseline SUV parameters for these lesions - and the subsequent 24-week response (Mann-Whitney U tests – see Figure 3-2), although this highest SUV parameters were identified in lesions with PD.

Response analysis of just the breast lesions also shows no association between the baseline SUV and 24-week treatment response (Mann-Whitney U tests: SUV_{mean} p=1.0; SUV_{max} p=0.752; SUV_{peak} p=0.709).

The baseline SUV parameters have no utility for predicting a treatment response in these patients.

3.1.1.2 Volumetric Parameters

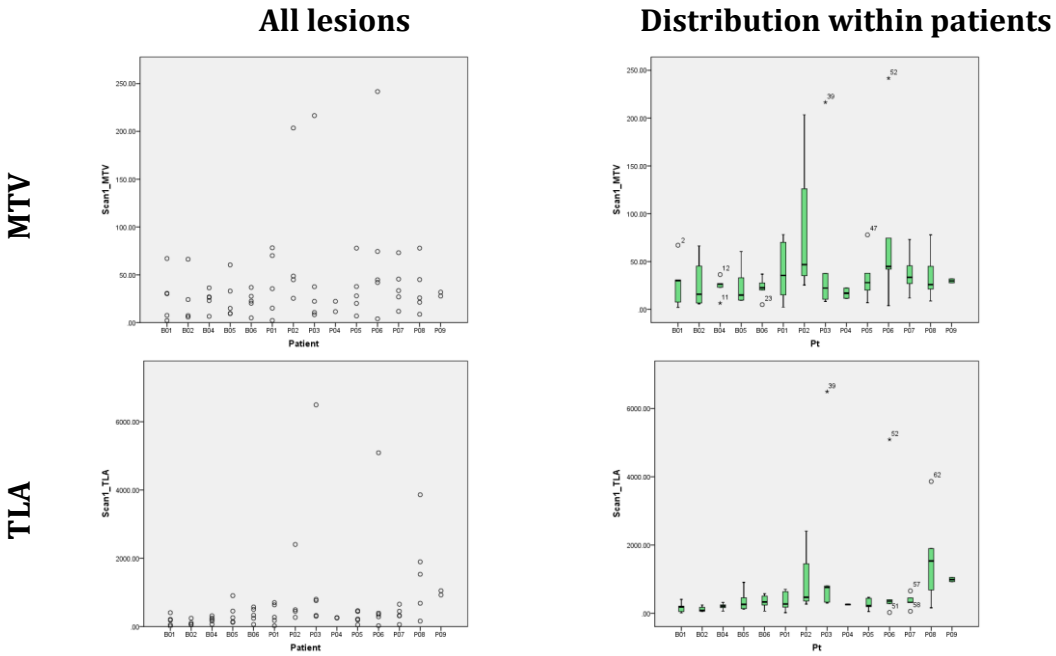


Figure 3-3: ^{18}F-Fluoride Per-lesion analysis; Baseline Volumetric Parameters

Fig.3-3 shows the range of volumetric parameters identified; the prostate patients yielded the largest measurements. The population data and tumour group data is recorded Tbl.3-2 and Fig.3-4.

	MTV	TLA
	Mean (95% CI) [min-max] {median; SD} |Skewness; Kurtosis|	
All lesions n=62	**39.7 (27.94-51.46)** [2.25-241.7] {27.29; 46.32} |3.04; 10.18|	**647.17 (359-935.35)** [15.77-6494.17] {310.79; 1134.76} |3.77; 15.18|
Breast (B) n=24	**25.04 (17.17-32.91)** [2.25-67.09] {23.62; 18.63} |1.03; 0.61|	**249.6 (162.76-336.45)** [15.77-907.08] {210.55; 205.67} |1.62; 3.34|
Prostate (P) n=38 **(7 did not have response analysis data)**	**48.96 (30.67-67.24)** [2.35-241.7] {32.68; 55.63} |2.47; 5.94|	**898.27 (441.71-1354.82)** [22.86-6494.17] {421.55; 1389.02} |2.91; 8.51|
Progression n=39 (B=10 P=29)	**47.21 (29.12-65.31)** [2.25-241.7] {30.71; 55.81} |2.45; 5.91|	**835.83 (385.79-1285.87)** [15.77-6494.17] {357.08; 1388.31} |2.95; 8.77|
Response n=16 (B=14, P=2)	**23.08 (14.87-31.3)** [4.96-66.31] {22.82; 15.42} |1.35; 3.17|	**233.45 (152.81-314.08)** [59.93-575.12] {244.38; 151.32} |0.88; 0.6|

Table 3-2: ^{18}F-Fluoride PET Per-lesion analysis; Descriptive statistics of baseline scan; Volumetric Parameters

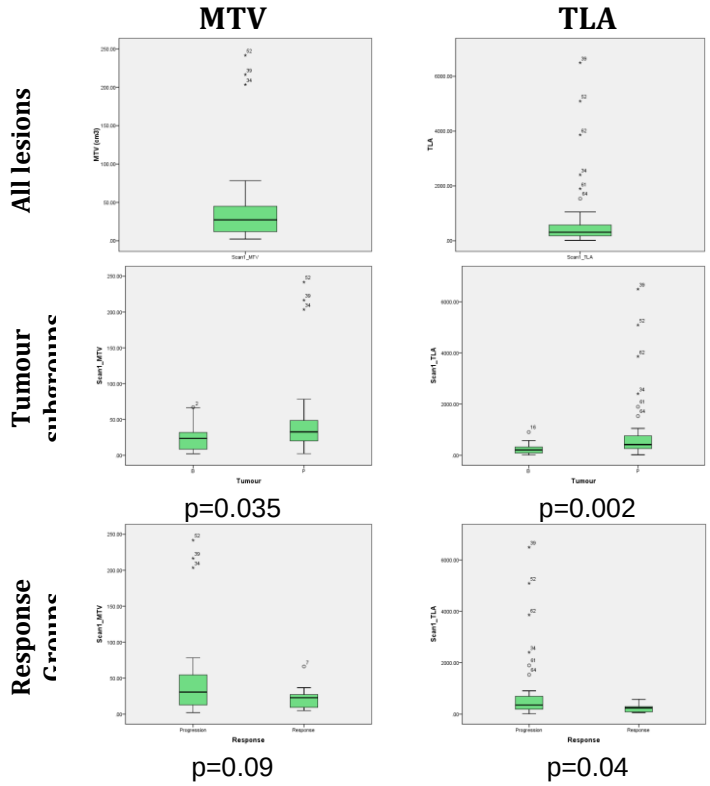


Figure 3-4: ^{18}F-Fluoride PET Per-lesion analysis; Descriptive statistics of baseline scan; Volumetric Parameters (B - Breast cancer patients; P – Prostate cancer)

There is a demonstrable difference in the distribution of MTV and TLA between the tumour groups, with larger lesions and higher TLA identified in the prostate cancer patients (Mann Whitney U tests: MTV p=0.035; TLA p=0.002).

There is a significant and expected correlation between the TLA and the MTV (Spearman's rank analysis: ρ=0.749, p<0.00001); TLA is the product of the MTV and the SUV_{mean}.

There is a notable difference in the distribution of the baseline volumetric parameters between response groups, with lesions in patients with PD having a higher MTV (p=0.09) and TLA (p=0.04). However, analysis of the breast cancer lesions alone does not identify an (Mann Whitney U tests: MTV p=0.752; TLA p=0.977). Prostate cancer lesions are larger and were more likely to have PD.

3.1.1.3 Heterogeneity Parameters

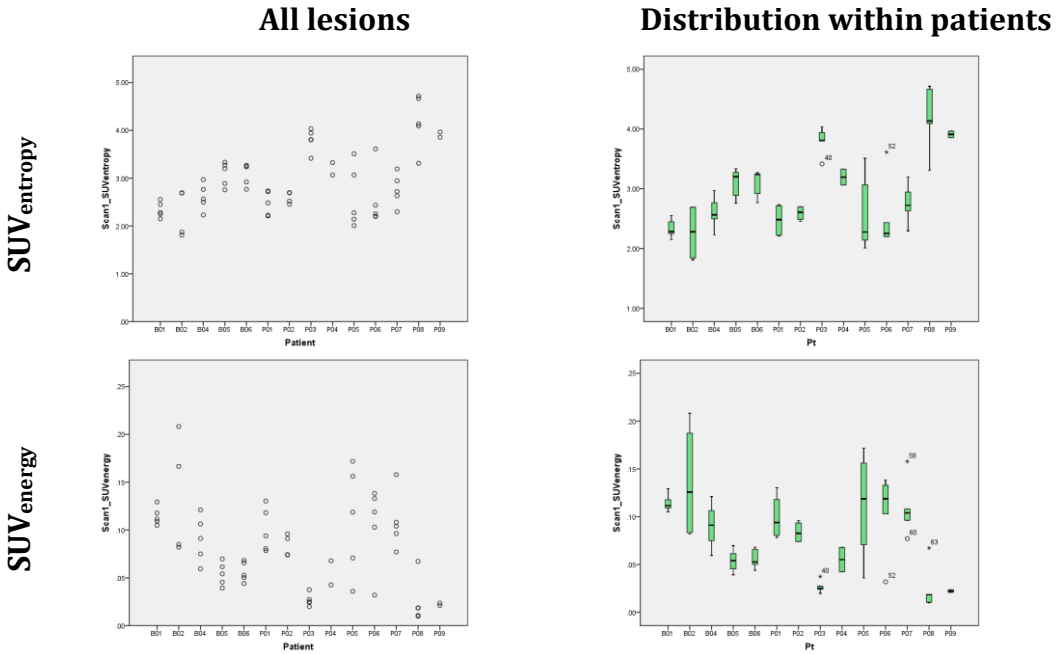


Figure 3-5: ^{18}F-Fluoride Per-lesion analysis; Baseline SUV Heterogeneity Parameters

	SUV$_{entropy}$	**SUV$_{energy}$**
	Mean (95% CI) [min-max] {median; SD} |Skewness; Kurtosis|	
All lesions n=62	**2.92 (2.75-3.09)** [1.81-4.72] {2.75; 0.68} |0.72; -0.02|	**0.0805 (0.0692-0.0918)** [0.0098-0.2082] {0.0761; 0.0446} |0.4793; -0.0786|
Breast n=24	**2.7 (2.51-2.88)** [1.81-3.34] {2.73; 0.45} |-0.29; -0.7|	**0.0883 (0.0709-0.1056)** [0.0393-0.2082] {0.0785; 0.0411} |1.2795; 1.902|
Prostate n=38	**3.06 (2.81-3.31)** [2.01-4.72] {2.84; 0.76} |0.51; -0.84|	**0.0756 (0.0603-0.0909)** [0.0098-0.1719] {0.0757; 0.0465} |0.255; -0.9813|
Progression n=39	**2.96 (2.72-3.21)** [2.01-4.72] {2.72; 0.75} |0.75; -0.47|	**0.0774 (0.063-0.0917)** [0.0098-0.1719] {0.0742; 0.0444} |0.149; -1.0034|
Response n=16	**2.75 (2.5-3)** [1.81-3.33] {2.77; 0.47} |-0.75; -0.1|	**0.0866 (0.0624-0.1109)** [0.0425-0.2082] {0.0716; 0.0455} |1.6685; 2.6197|

Table 3-3: ^{18}F-Fluoride PET Per-lesion analysis; Descriptive statistics of baseline scan; SUV Heterogeneity Parameters

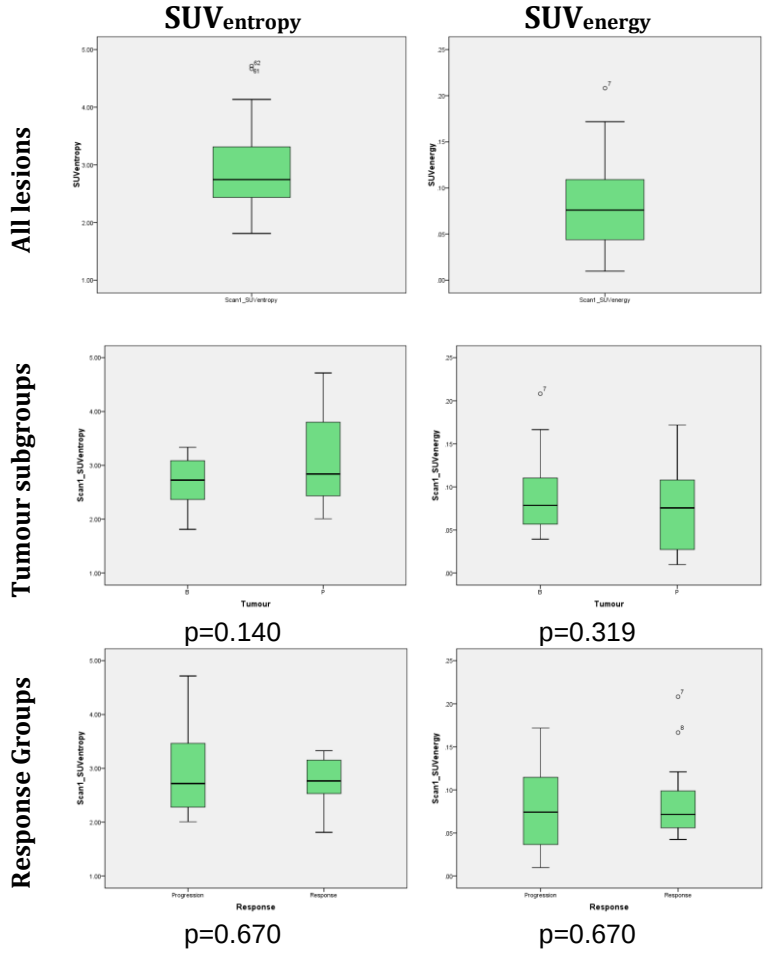


Figure 3-6: ^{18}F-Fluoride PET Per-lesion analysis; Descriptive statistics of baseline scan; SUV Heterogeneity Parameters (B - Breast cancer patients; P – Prostate cancer)

There is no demonstrable difference in the distribution of the baseline SUV heterogeneity parameters between the tumour groups (Mann-Whitney U tests: $SUV_{entropy}$ p=0.140 and for SUV_{energy} p=0.319), or response groups (Mann-Whitney U tests: $SUV_{entropy}$ p=0.670 and for SUV_{energy} p=0.0670). The highest $SUV_{entropy}$ and lowest SUV_{energy} measurements were in lesions representing PD. Analysis of the breast cancer lesions only also demonstrates no predictive utility (p=1.0 for both parameters).

3.1.1.4 Ki

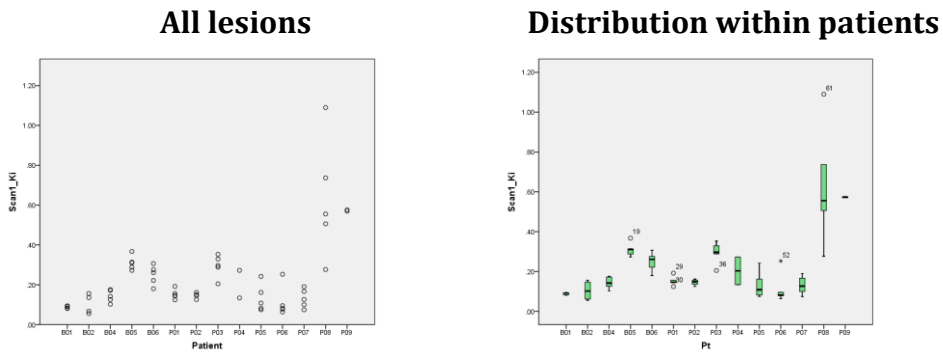


Figure 3-7: ^{18}F-Fluoride Per-lesion analysis; Baseline Ki

	Ki
	Mean (95% CI) [min-max] {median; SD} |Skewness; Kurtosis|
All lesions n=62	**0.22 (0.18-0.27)** [0.06-1.09] {0.16; 0.18} |2.54; 8.57|
Breast n=24	**0.18 (0.14-0.22)** [0.06-0.37] {0.16; 0.09} |0.41; -1.22|
Prostate n=38	**0.25 (0.18-0.32)** [0.06-1.09] {0.16; 0.22} |2.2; 5.57|
Progression n=39	**0.24 (0.17-0.3)** [0.06-1.09] {0.16; 0.2} |2.5; 7.9|
Response n=16	**0.17 (0.13-0.21)** [0.06-0.31] {0.16; 0.08} |0.28; -0.79|

Table 3-4: ^{18}F-Fluoride PET Per-lesion analysis; Descriptive statistics of baseline scan; Ki

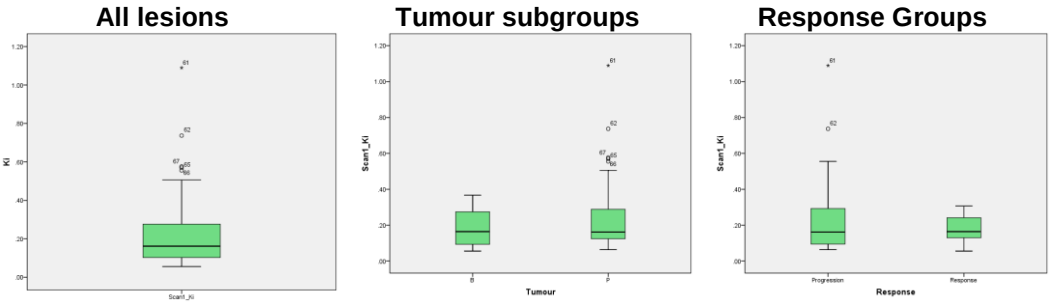


Figure 3-8: ^{18}F-Fluoride PET Per-lesion analysis; Descriptive statistics of baseline scan; Ki (B - Breast cancer patients; P – Prostate cancer)

Fig.3-7 and Tbl.3-4 report the data of the baseline Ki parameter, showing the inter-lesion heterogeneity. There is no demonstrable statistically significant difference in the distribution of Ki between tumour subgroups (Mann-Whitney U test: p=0.479), or 24-week response groups (Mann-Whitney U test, p=0.541). There is no significant difference between the response groups when only the breast cancer lesions are analysed (p=0.709).

Correlations between parameters

	SUV_{mean}	SUV_{max}	SUV_{peak}	MTV	TLA	$SUV_{entropy}$	SUV_{energy}	Ki
SUV_{mean}	1	-	-	-	-	-	-	-
SUV_{max}	0.843 (<0.0001)	1	-	-	-	-	-	-
SUV_{peak}	0.802 (<0.0001)	0.983 (<0.0001)	1	-	-	-	-	-
MTV	-0.07 (0.59)	0.218 (0.09)	0.17 (0.19)	1	-	-	-	-
TLA	0.549 (<0.0001)	0.679 (<0.0001)	0.619 (<0.0001)	0.749 (<0.0001)	1	-	-	-
$SUV_{entropy}$	0.957 (<0.0001)	0.891 (<0.0001)	0.865 (<0.0001)	-0.021 (0.87)	0.572 (<0.0001)	1	-	-
SUV_{energy}	-0.965 (<0.0001)	-0.821 (<0.0001)	-0.779 (<0.0001)	0.018 (0.89)	-0.581 (<0.0001)	-0.973 (<0.0001)	1	-
Ki	0.968 (<0.0001)	0.802 (<0.0001)	0.759 (<0.0001)	-0.11 (0.39)	0.491 (<0.0001)	0.927 (<0.0001)	-0.944 (<0.0001)	1

Table 3-5: ^{18}F-Fluoride PET Per-lesion analysis; Baseline Parameters; Correlations between parameters (Spearman's Correlation coefficient (p-value))

To complicate the analysis, there are significant correlations between most of the baseline parameters (Tbl.3-5).

The correlation between the SUV parameters was anticipated because they are all describing the same histogram. TLA is the product of MTV and SUV_{mean}, and therefore the correlation between TLA and the SUV parameters is also to be expected. MTV is not significantly correlated with any of the other parameters (other than TLA). The heterogeneity parameters are significantly and strongly correlated with the SUV parameters; $SUV_{entropy}$ is positively correlated with the SUV parameters, and SUV_{energy} negatively correlated. Ki is significantly and strongly correlated with all the parameters other than MTV. MTV seems to be the most independent parameter used in this study.

3.1.2 Second Scan

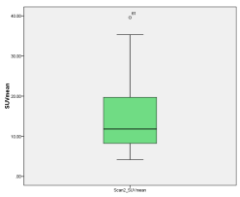
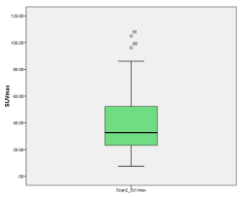
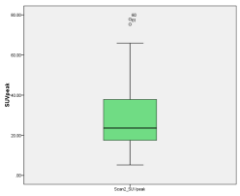
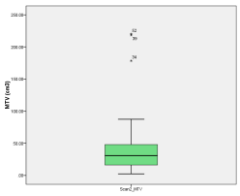
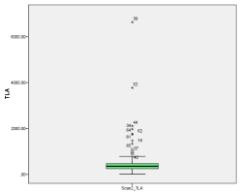
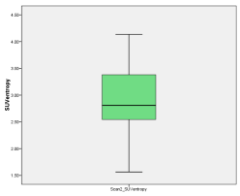
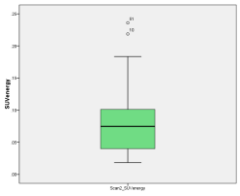
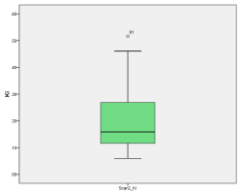
	Mean (95% CI) [min-max] {median; SD} |Skewness; Kurtosis|	**Box-Plots**
SUV_{mean}	**14.95 (12.53-17.37)** [4.18-39.57] {11.8; 9.21} |1.12; 0.26|	
SUV_{max}	**39.48 (33.41-45.56)** [7.66-104.96] {32.73; 23.1} |1.02; 0.31|	
SUV_{peak}	**30.05 (25.37-34.73)** [5.21-77.8] {23.59; 17.8} |1.03; 0.22|	
MTV	**40.7 (29.23-52.17)** [2.4-220.09] {30.71; 43.62} |3; 9.85|	
TLA	**647.7 (377.11-918.29**) [11.64-6630.26] {356.53; 1029.1} |4.18; 20.98|	
$SUV_{entropy}$	**2.93 (2.77-3.09)** [1.56-4.13] {2.81; 0.62} |0.15; -0.54|	
SUV_{energy}	**0.0792 (0.0662-0.0922)** [0.0184-0.2359] {0.0747; 0.0494} |1.1739; 1.4403|	
Ki	**0.21 (0.17-0.24)** [0.06-0.52] {0.16; 0.12} |0.9; -0.22|	

Table 3-6: ^{18}F-Fluoride PET Per-lesion analysis; Descriptive statistics of second scan; All Parameters

3.1.3 %Δ Parameters

3.1.3.1 %ΔSUV Parameters

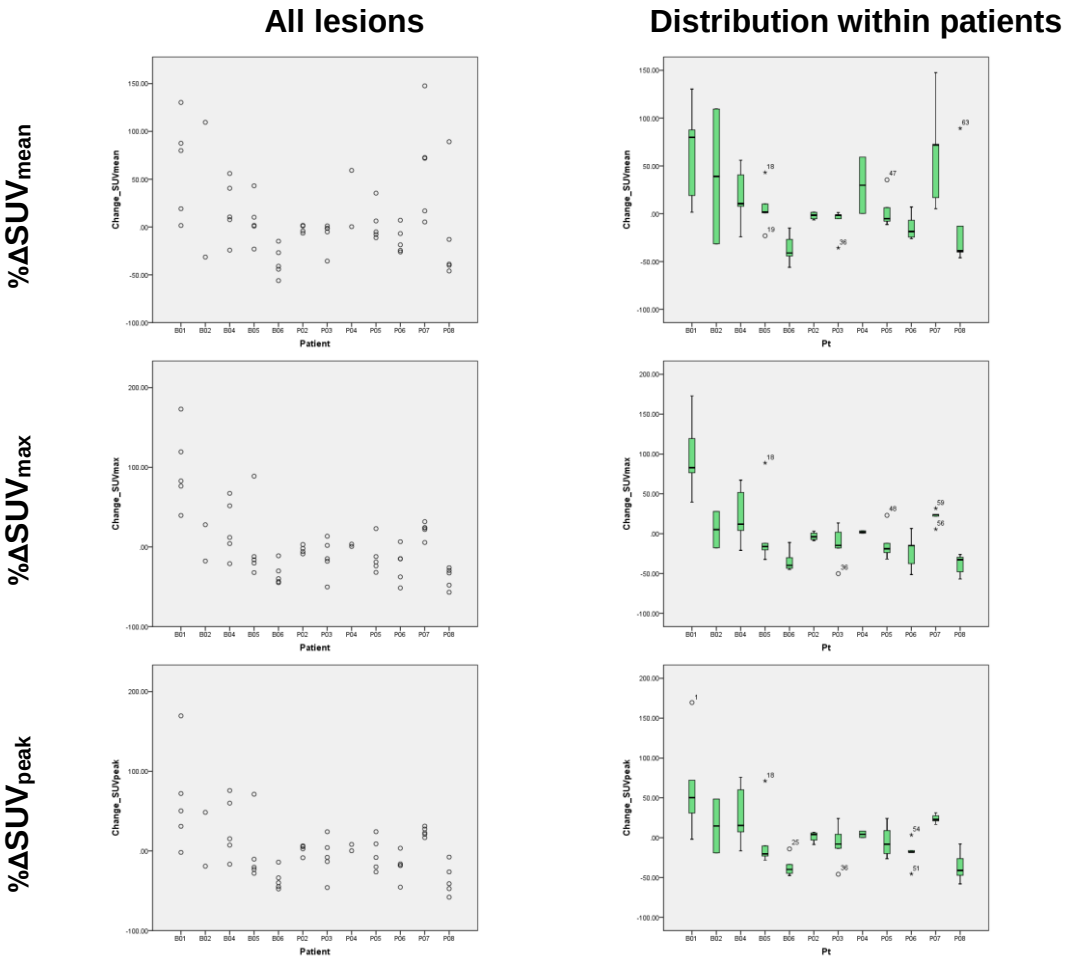


Figure 3-9: ^{18}F-Fluoride Per-lesion analysis; %Δ SUV Parameters

	%ΔSUV$_{mean}$	%ΔSUV$_{max}$	%ΔSUV$_{peak}$
		Mean (95% CI) [min-max] {median; SD} |Skewness; Kurtosis|	
All lesions n=53	**10.59 (-1.98-23.16)** [-56.01-147.59] {0.92; 45.61} |1.21; 1.14|	**2.2 (-10.16-14.57)** [-56.87-173] {-11.07; 44.87} |1.64; 3.5|	**1.89 (-9.12-12.9)** [-57.97-169.57] {-7.67; 39.95} |1.63; 4.82|
Breast (B) n=22	**15.33 (-7.35-38.01)** [-56.01-130.28] {4.74; 51.15} |0.79; -0.11|	**20.66 (-5.7-47.01)** [-44.84-173] {-3.44; 59.44} |1; 0.38|	**13.71 (-10-37.42)** [-47.58-169.57] {-6.06; 53.47} |1.25; 1.82|
Prostate (P) n=31	**7.23 (-8.1-22.56)** [-46.01-147.59] {-1.5; 41.79} |1.7; 3.37|	**-10.89 (-19.79--1.99)** [-56.87-31.76] {-12.05; 24.27} |-0.14; -0.74|	**-6.5 (-15.41-2.4)** [-57.97-31.25] {-7.67; 24.28} |-0.41; -0.6|
Progression n=34 B=10 P=24	**5.93 (-7.74-19.6)** [-46.01-130.28] {-1.41; 39.18} |1.61; 2.67|	**1.48 (-16.41-19.37)** [-56.87-173] {-14.72; 51.27} |1.79; 3.21|	**-0.55 (-15.47-14.37)** [-57.97-169.57] {-8.35; 42.75} |2.15; 6.82|
Response n=14 B=12 P=2	**3.24 (-24.3-30.78)** [-56.01-109.53] {-7.26; 47.7} |0.89; 0.21|	**-2.92 (-22.72-16.89)** [-44.84-67.29] {-5.21; 34.3} |0.72; -0.09|	**-0.03 (-22.62-22.56)** [-47.58-75.85] {-6.93; 39.12} |0.69; -0.45|

Table 3-7: ^{18}F-Fluoride PET Per-lesion analysis; Descriptive statistics of %ΔSUV Parameters

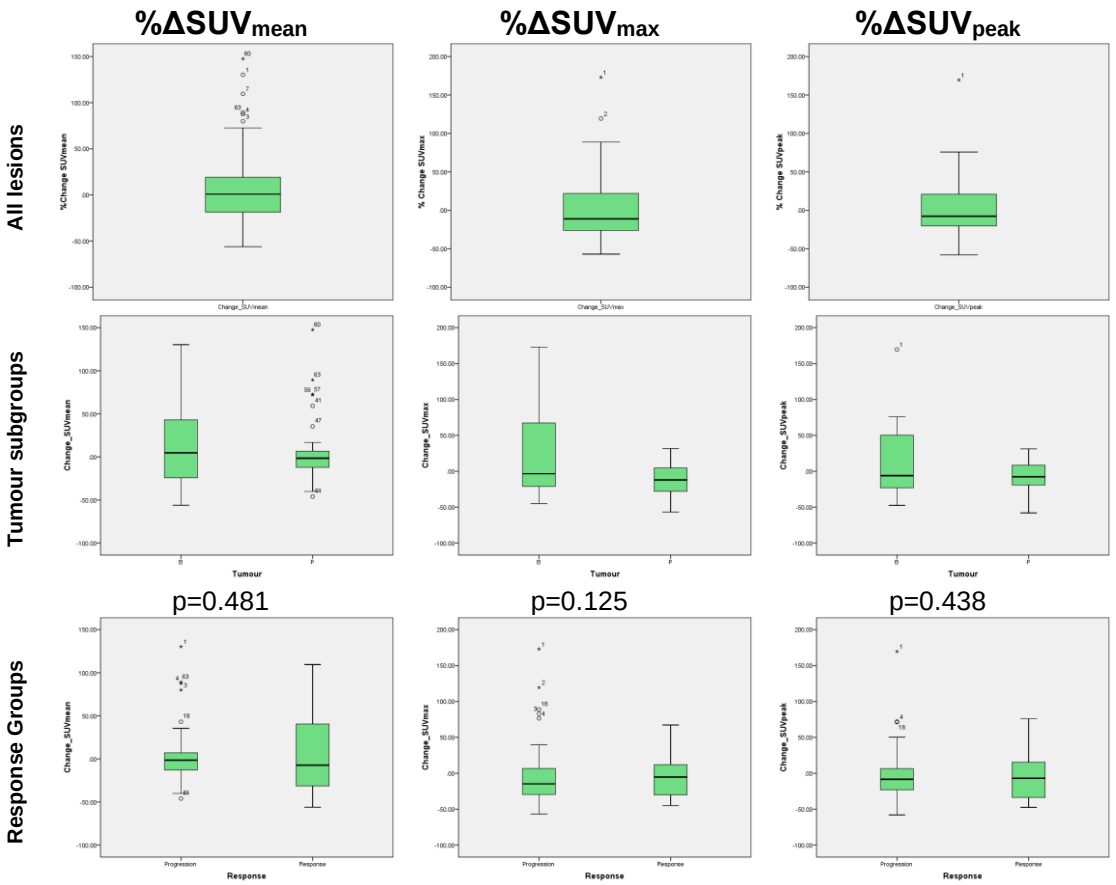


Figure 3-10: ^{18}F-Fluoride PET Per-lesion analysis; Descriptive statistics of %ΔSUV Parameters (B - Breast cancer patients; P – Prostate cancer)

Fig.3-9, Fig.3-10, and Tbl.3-7 show lesions with both increases and decrease in the %ΔSUV parameters have been identified. Within individual patients there are lesions with opposing changes.

There is no statistically significant difference in the distribution of the %ΔSUV Parameters between the response groups (Fig.3-24). Within each response group there are lesions for which the parameter has increased and for which the parameter has fallen.

Analysis of only the breast cancer lesions demonstrates a trend towards significance for identification of the 24-week response group from the %Δ SUV parameters (Fig.3-11)).

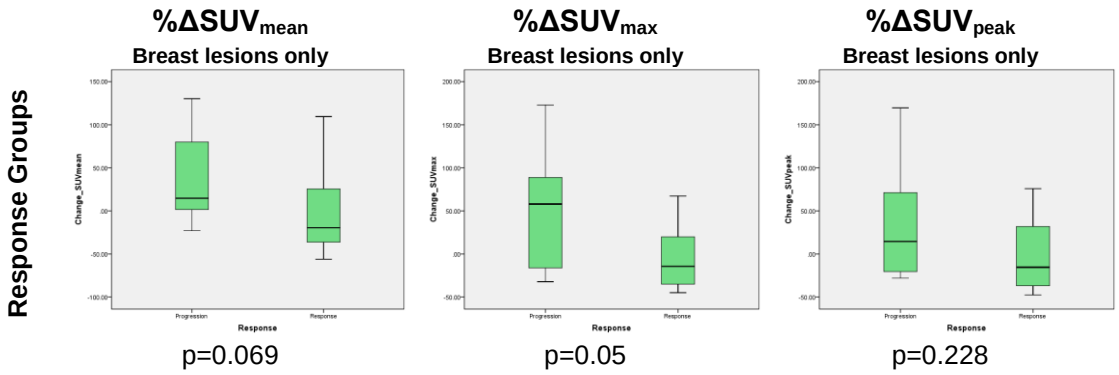
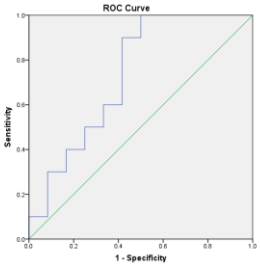
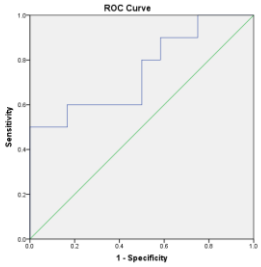


Figure 3-11: ^{18}F-Fluoride PET Per-lesion analysis; Descriptive statistics of %ΔSUV Parameters; Breast lesions only; comparisons with Mann-Whitney U tests

Statistical significance is only reached at the 0.05 level for %ΔSUV$_{max}$. The %ΔSUV$_{mean}$ data approaches significance (p=0.069), but although the box-plots suggest a similar separation of the %ΔSUV$_{peak}$ data, no statistically significant difference has been identified (p=0.228).

The ROC data (Fig.3-12) can be used to identify the accuracy shown by these two parameters for the prediction of PD by 24-weeks.

	%ΔSUV$_{mean}$ Breast lesions only	%ΔSUV$_{max}$ Breast lesions only
ROC		
AUC	0.733	0.750
p-value	0.065	0.048

Parameter	Parameter value	Sensitivity	Specificity	AUC	p-value
%ΔSUV$_{mean}$	>+1.25%	80%	58%	0.733	0.065
%ΔSUV$_{max}$	>+19.91%	60%	75%	0.775	0.048

Figure 3-12: ^{18}F-Fluoride PET Per-lesion analysis; Response analysis of %Δ ADC parameters with ROC analysis for prediction of PD; Breast lesions only

The %Δ in SUV$_{mean}$ shows increased sensitivity for predicting PD, but the %Δ SUV$_{max}$ shows a higher specificity.

Despite this association, there were lesions within each response group showing inconsistency in the direction of SUV parameter change. The heterogeneous change in these parameters can be visualised using waterfall plots (Fig.3-13).

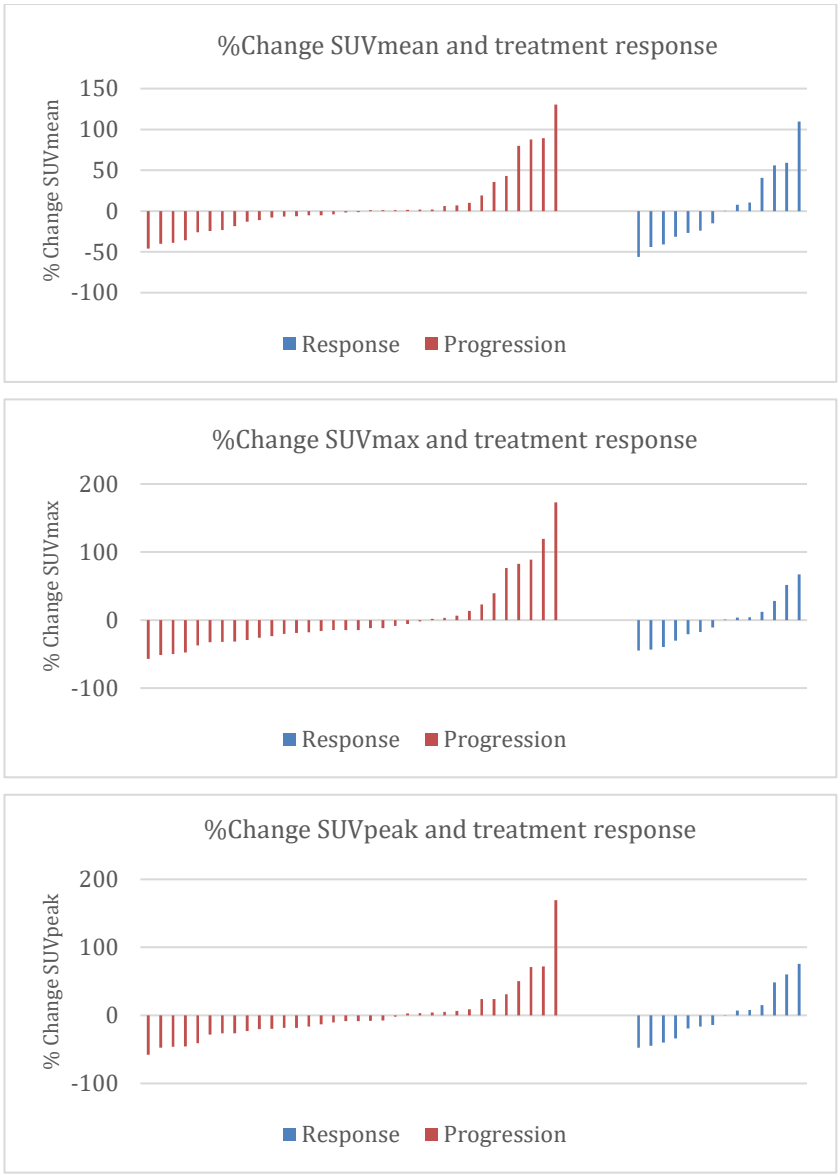


Figure 3-13: ^{18}F-Fluoride PET Per-lesion analysis; Response analysis of %Δ SUV Parameters; Waterfall plots for all lesions

3.1.3.2 %Δ Volumetric Parameters

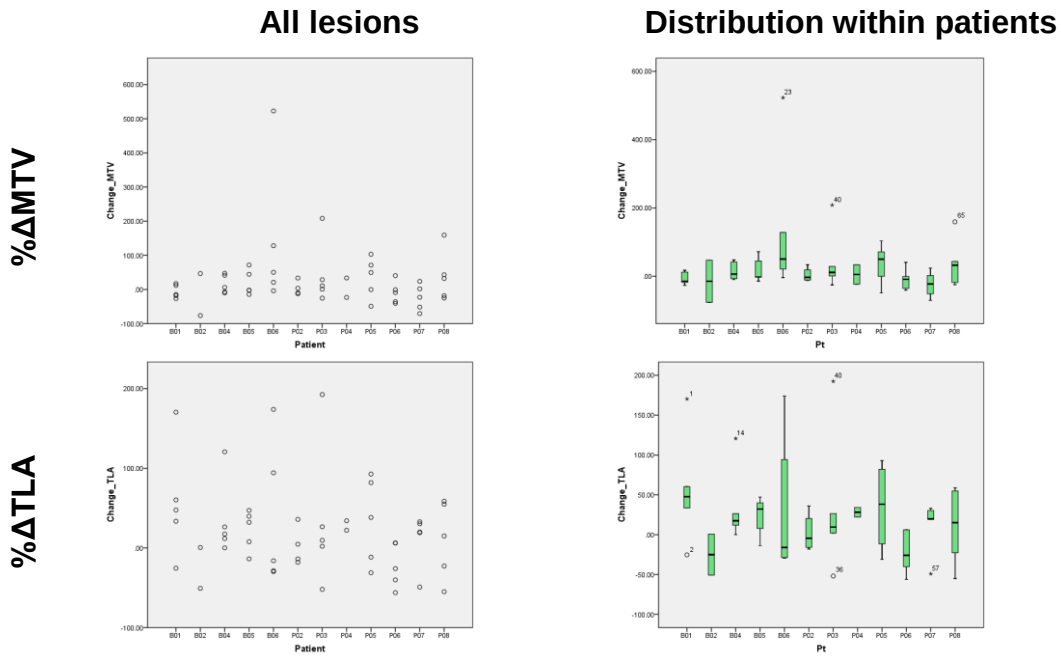
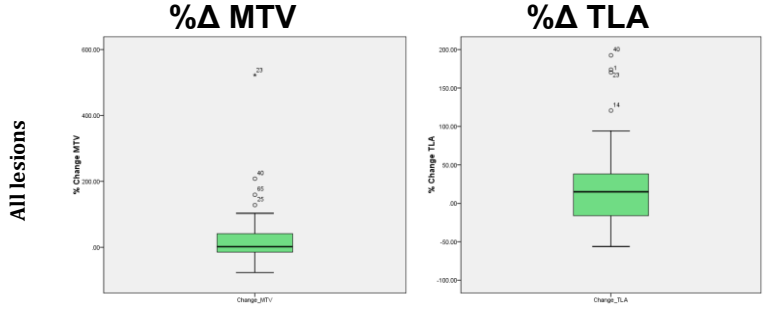


Figure 3-14: ^{18}F-Fluoride Per-lesion analysis; %Δ Volumetric Parameters

	%ΔMTV	%ΔTLA
	Mean (95% CI) [min-max] {median; SD} |Skewness; Kurtosis|	
All lesions n=53	**24.24 (0.25-48.23)** [-76.47-522.66] {1.98; 87.03} |3.94; 20.78|	**21.3 (6.02-36.58)** [-56.09-192.47] {15.04; 55.45} |1.27; 2.01|
Breast (B) n=22	**37.94 (-13.46-89.33)** [-76.47-522.66] {9.18; 115.92} |3.77; 16.01|	**32.73 (5.78-59.67)** [-50.71-173.9] {21.91; 60.77} |1.12; 0.85|
Prostate (P) n=31	**14.52 (-7.14-36.19)** [-70.51-208.24] {0.86; 59.05} |1.64; 3.49|	**13.19 (-5.44-31.82)** [-56.09-192.47] {9.56; 50.79} |1.42; 4.01|
Progression n=34 B=10 P=24	**18.43 (-0.61-37.47)** [-49.12-208.24] {0.25; 54.57} |1.86; 4.18|	**20.56 (1.03-40.09)** [-56.09-192.47] {8.7; 55.96} |1.32; 2.49|
Response n=14 B=12 P=2	**55.51 (-26.66-137.68)** [-76.47-522.66] {27.34; 142.32} |3.07; 10.49|	**26.9 (-9.22-63.02)** [-50.71-173.9] {14.63; 62.56} |1.23; 1.15|

Table 3-8: ^{18}F-Fluoride PET Per-lesion analysis; Descriptive statistics of %Δ Volumetric Parameters



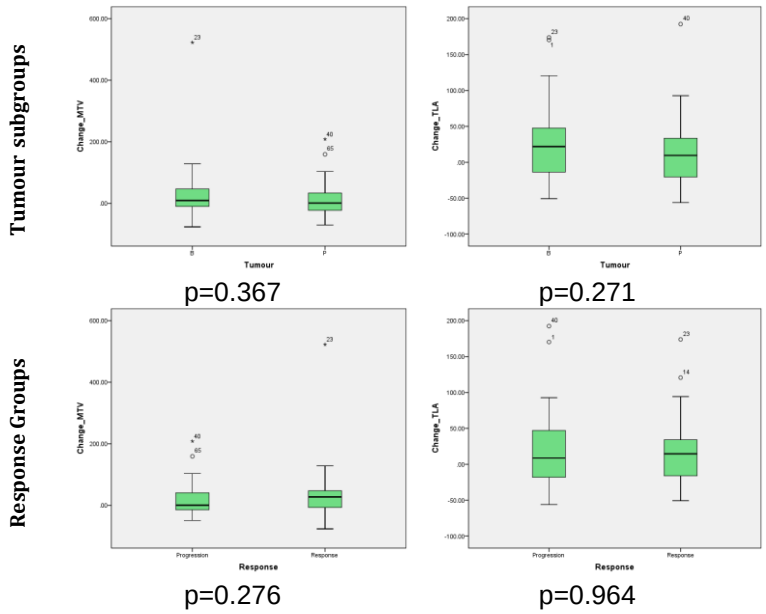


Figure 3-15: ^{18}F-Fluoride PET Per-lesion analysis; Descriptive statistics of %Δ Volumetric Parameters (B - Breast cancer patients; P – Prostate cancer)

There is a wide spread of results, even within individual patients (Fig.3-14). Tbl.3-8 and Fig.3-15 report this data and compares the tumour groups and response groups.

There is no demonstrable statistically significant difference between the distributions of the %Δ volumetric parameters and the response group (Mann-Whitney U Tests: %ΔMTV p=0.367; %ΔTLA p=0.271). There is also no difference when only the breast cancer lesions are analysed (Mann-Whitney U Tests: %ΔMTV p=0.180; %ΔTLA p=0.254). Both response groups contain lesions with increases and decreases in volumetric parameter (Fig.3-16).

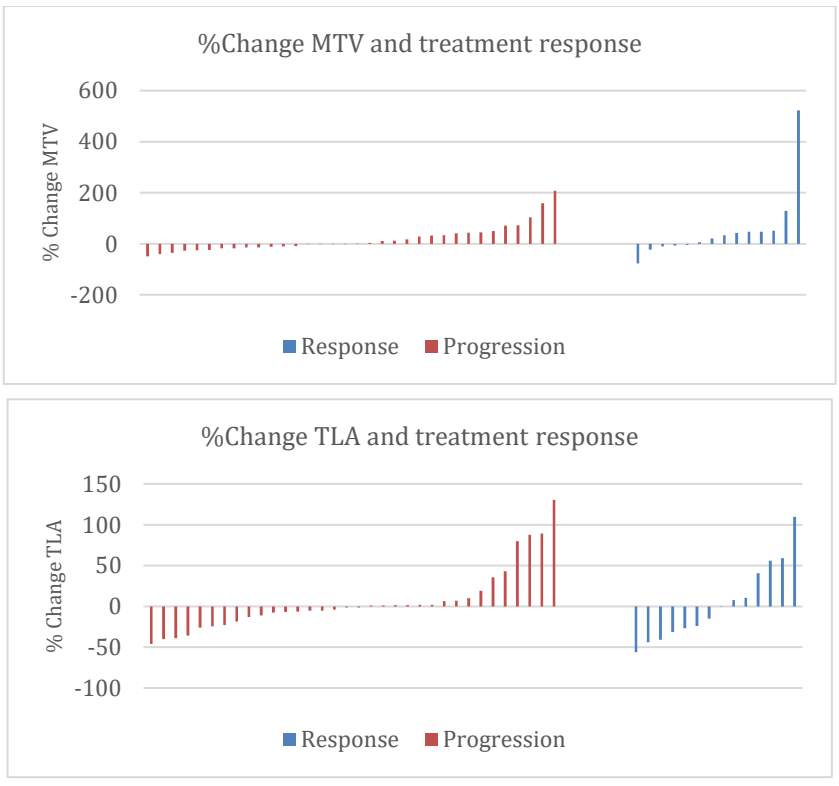


Figure 3-16: ^{18}F-Fluoride PET Per-lesion analysis; Response analysis of %Δ Volumetric Parameters; Waterfall plots for all lesions

3.1.3.3 %Δ Heterogeneity Parameters

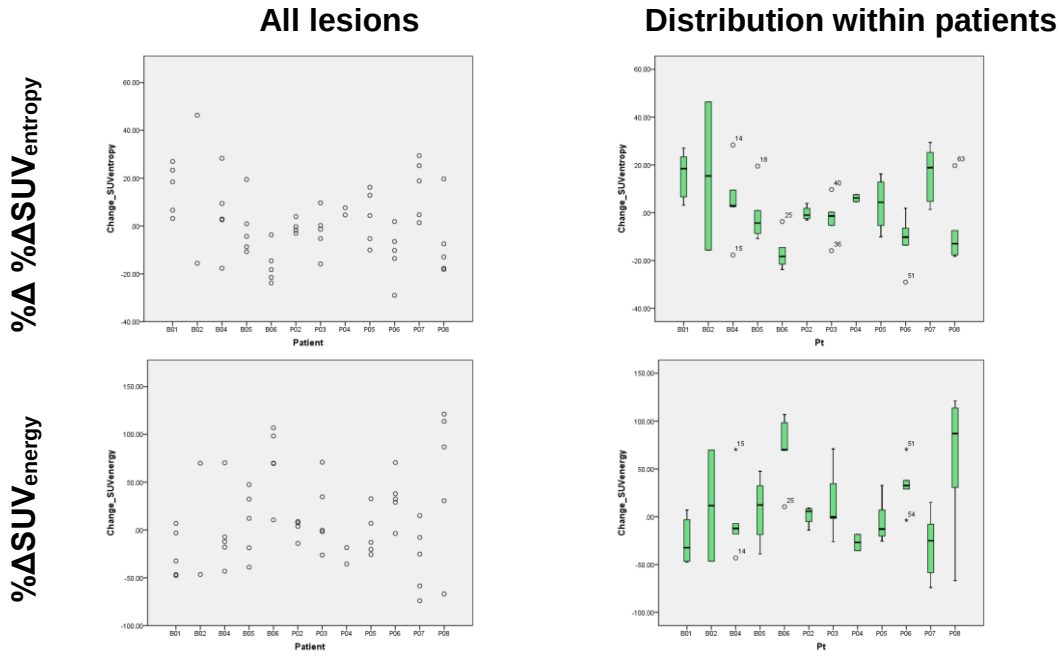


Figure 3-17: ^{18}F-Fluoride Per-lesion analysis; %Δ SUV Heterogeneity Parameters

	%ΔSUV$_{entropy}$	%ΔSUV$_{energy}$
	Mean (95% CI) [min-max] {median; SD} |Skewness; Kurtosis|	
All lesions n=53	**0.99 (-3.33-5.31)** [-29-46.4] {0.26; 15.67} |0.54; 0.17|	**11.19 (-2.02-24.39)** [-74-121.23] {3.84; 47.92} |0.53; -0.37|
Breast (B) n=22	**2.27 (-6.03-10.56**) [-23.81-46.4] {1.76; 18.71} |0.61; -0.2|	**12.73 (-9.45-34.91)** [-47.65-106.89] {1.92; 50.02} |0.47; -1.07|
Prostate (P) n=31	**0.08 (-4.83-4.98)** [-29-29.4] {-0.27; 13.37} |0.22; -0.01|	**10.09 (-7.22-27.39)** [-74-121.23] {3.84; 47.18} |0.6; 0.37|
Progression n=34 B=10 P=24	**-0.42 (-4.99-4.15)** [-29-27.01] {-1.55; 13.1} |0.27; -0.25|	**12.61 (-2.75-27.97)** [-66.83-121.23] {6.96; 44.03} |0.69; 0.42|
Response n=14 B=12 P=2	**-0.94 (-12.56-10.68)** [-23.81-46.4] {-0.55; 20.13} |1.07; 0.98|	**22.45 (-9.51-54.4)** [-46.5-106.89] {1.64; 55.34} |0.24; -1.7|

Table 3-9: ^{18}F-Fluoride PET Per-lesion analysis; Descriptive statistics of %Δ SUV Heterogeneity Parameters

As for the other %Δ parameters, a full range of change has been identified in these study lesions, and there is again a heterogeneous change within individual patients (Fig.3-17). Table 3-9 reports the full data. There is no demonstrable difference identified in the distribution of the %Δ heterogeneity parameters between the tumour subgroups (Fig.3-18).

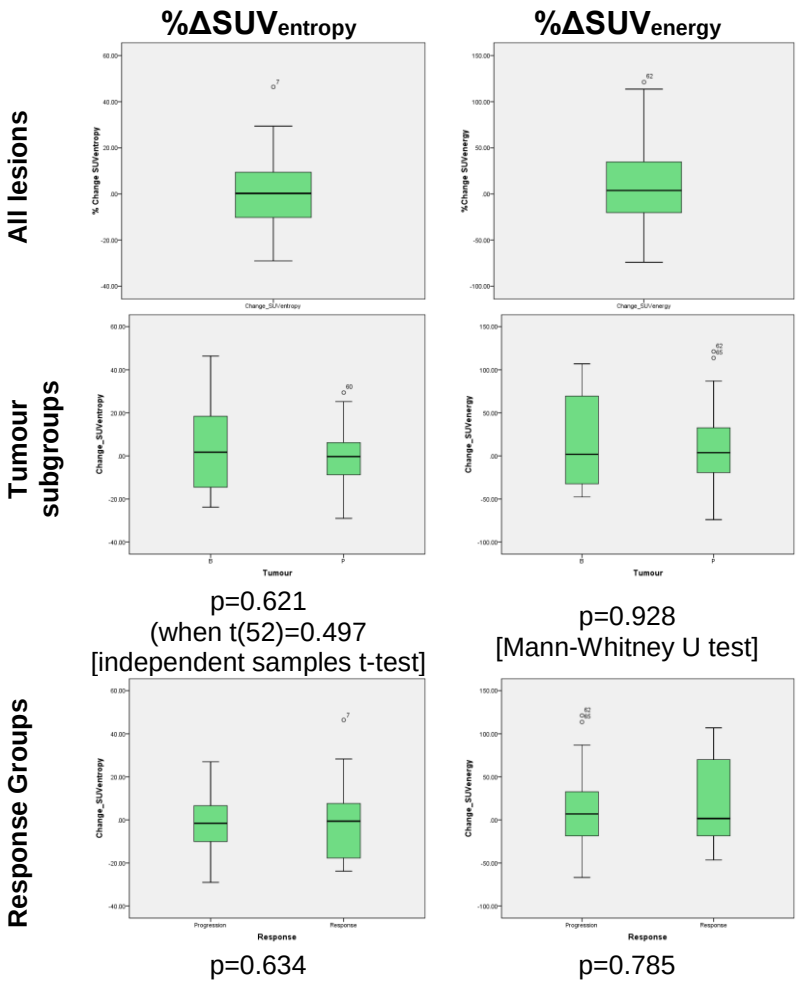


Figure 3-18: ^{18}F-Fluoride PET Per-lesion analysis; Descriptive statistics of %Δ SUV Heterogeneity Parameters (B - Breast cancer patients; P – Prostate cancer)

There is no statistically significant difference in the distribution of the %Δ heterogeneity parameters between the 24-week response groups for these lesions (Tbl.3-18). Analysis of just the breast cancer lesions can be reviewed below (Fig.3-19).

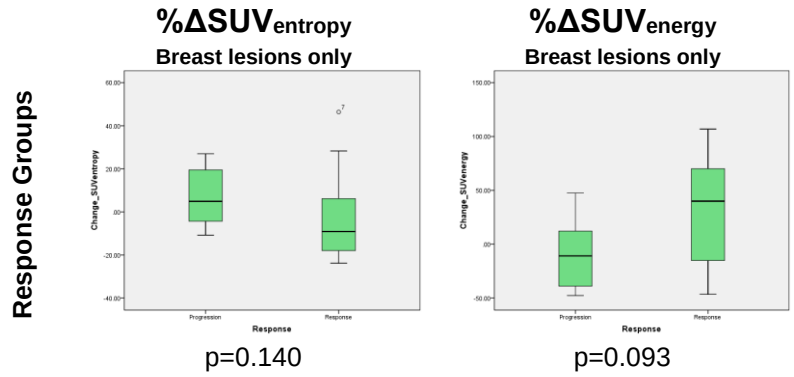


Figure 3-19: ^{18}F-Fluoride PET Per-lesion analysis; Descriptive statistics of %Δ SUV Heterogeneity Parameters; breast lesions only; comparisons with Mann-Whitney U tests

There is a trend towards significance for PD being predicted for the breast cancer lesions with an increase in $SUV_{entropy}$ between scans (p=0.140) and a decrease in SUV_{energy} (p=0.093), i.e. when treatment results in a more heterogeneous SUV voxel distribution.

Again, for both response groups there are lesions showing an increase in the heterogeneity parameter, but also lesions for which the heterogeneity parameter fell with treatment. This can be more clearly seen in the waterfall plots below:

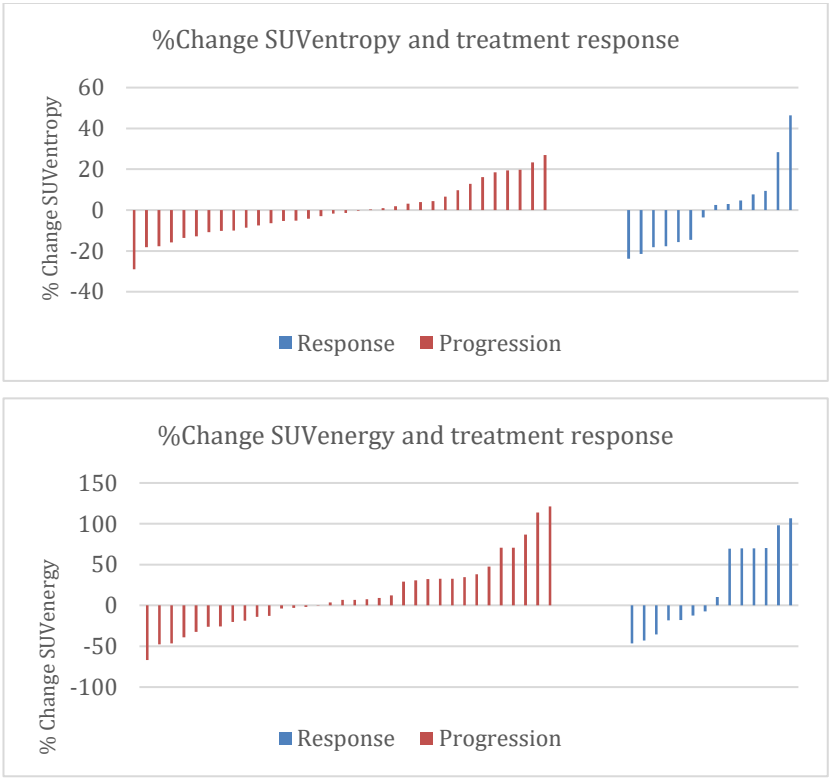


Figure 3-20: ^{18}F-Fluoride PET Per-lesion analysis; Response analysis of %Δ SUV heterogeneity parameters; Waterfall plots for all lesions

3.1.3.4 %Δ Ki

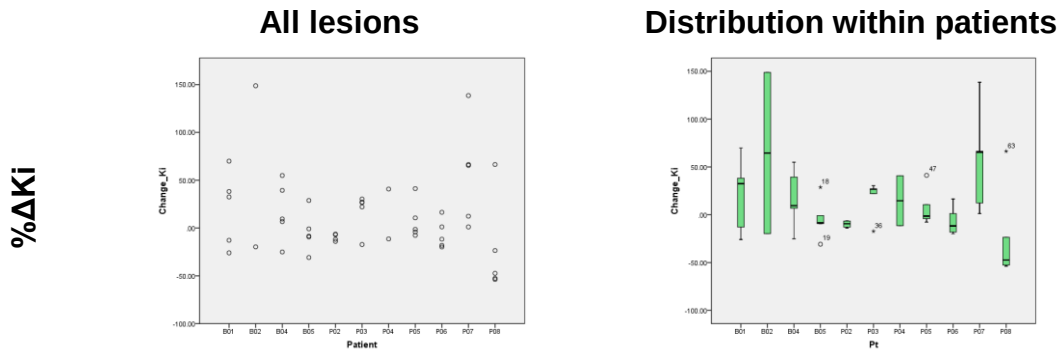


Figure 3-21: ^{18}F-Fluoride Per-lesion analysis; %Δ Ki

	%ΔKi
	Mean (95% CI) [min-max] {median; SD} |Skewness; Kurtosis|
All lesions n=48	**11.56 (-0.49-23.6)** [-53.77-148.76] {0.15; 41.49} |1.34; 2.68|
Breast n=17	**17.4 (-5.91-40.71)** [-30.83-148.76] {6.69; 45.34} |1.59; 3.32|
Prostate n=31	**8.35 (-6.18-22.89)** [-53.77-138.52] {-1.4; 39.63} |1.2; 2.7|
Progression n=34	**0.79 (-9.66-11.24)** [-53.77-69.96] {-6.66; 29.96} |0.4; 0.07|
Response n=9	**27.09 (-14.16-68.34)** [-25.07-148.76] {9.51; 53.66} |1.57; 3|

Table 3-10: ^{18}F-Fluoride PET Per-lesion analysis; Descriptive statistics of %Δ Ki

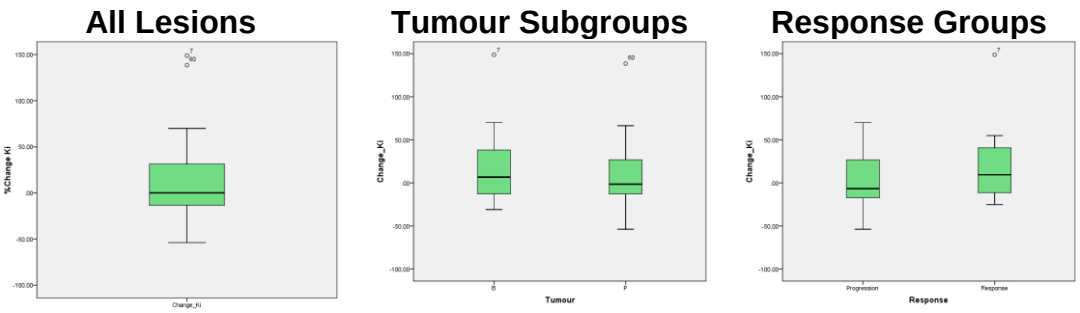


Figure 3-22: ^{18}F-Fluoride PET Per-lesion analysis; Descriptive statistics of %Δ Ki (B - Breast cancer patients; P – Prostate cancer)

As for the other parameters in this study, a full range of changes in Ki has been identified, including within individual patients (Fig.3-21). Table 3-10 reports the data.

Statistical comparison of the distributions of %ΔKi between the response group does not reach significance (Fig.3-22), but the Ki seemed higher in lesions from patients with a treatment response at 24-weeks. There is no statistically significant difference identified when only breast cancer lesions are analysed (p=0.417).

The heterogeneous change in parameter seen with the other parameters has also been identified in these lesions, with positive and negative changes seen in each response group. This can be seen clearly on the following waterfall plot:

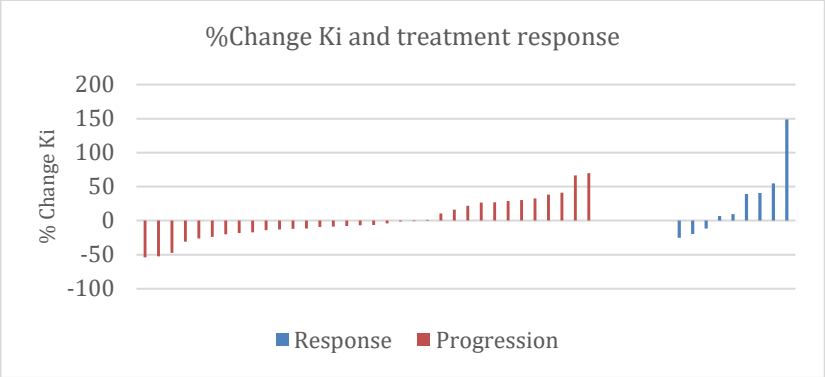


Figure 3-23: ^{18}F-Fluoride PET Per-lesion analysis; Response analysis of %ΔKi; Waterfall plots for all lesions

3.1.3.5 Correlations between %Δ parameters

	%ΔSUV$_{mean}$	%ΔSUV$_{max}$	%ΔSUV$_{peak}$	%ΔMTV	%ΔTLA	%ΔSUV$_{entropy}$	%ΔSUV$_{energy}$	%ΔKi
%ΔSUV$_{mean}$	1	-	-	-	-	-	-	-
%ΔSUV$_{max}$	0.738 (<0.0001)	1	-	-	-	-	-	-
%ΔSUV$_{peak}$	0.783 (<0.0001)	0.938 (<0.0001)	1	-	-	-	-	-
%ΔMTV	0.299 (0.03)	0.314 (0.02)	0.346 (0.01)	1	-	-	-	-
%ΔTLA	-0.456 (<0.0001)	-0.211 (0.13)	-0.215 (0.12)	0.583 (<0.0001)	1	-	-	-
%ΔSUV$_{entropy}$	0.917 (<0.0001)	0.801 (<0.0001)	0.852 (<0.0001)	0.403 (<0.0001)	-0.308 (0.02)	1	-	-
%ΔSUV$_{energy}$	-0.908 (<0.0001)	-0.716 (<0.0001)	-0.776 (<0.0001)	-0.418 (<0.0001)	0.311 (0.02)	-0.969 (<0.0001)	1	-
%ΔKi	0.968 (<0.0001)	0.802 (<0.0001)	0.759 (<0.0001)	-0.11 (0.39)	0.491 (<0.0001)	0.927 (<0.0001)	-0.944 (<0.0001)	1

Table 3-11: ^{18}F-Fluoride PET Per-lesion analysis; %Change Parameters; Correlations between parameters (Spearman's Correlation coefficient (p-value))

There are significant correlations between most of the %Δ parameters (Table 3-11). The SUV parameters are describing the same histogram and are closely correlated. TLA is the product of MTV and SUV$_{mean}$, and the correlation between TLA and the SUV$_{mean}$ parameters is also to be expected. There is a trend towards a weak association for the other SUV parameters.

MTV was not significantly correlated with the other parameters (apart from TLA) at the baseline scan, but the %Δ MTV parameter shows a weak but significant correlation with the other parameters. The heterogeneity parameters are significantly and strongly correlated with the SUV parameters; SUV$_{entropy}$ is positively correlated with the SUV parameters, and SUV$_{energy}$ negatively correlated. Ki is significantly and strongly correlated with all the parameters other than MTV.

%Δ MTV still seems to be the most independent parameter used in this study, but still shows some weak correlations with the other parameters.

3.2 Discussion

		SUV_{mean}	SUV_{max}	SUV_{peak}	MTV	TLA	$SUV_{entropy}$	SUV_{energy}	Ki
Baseline	**Tumour**	↑ Prostate p=0.066	↑ Prostate p=0.011	↑ Prostate p=0.008	↑ Prostate p=0.035	↑ Prostate p=0.002	X	X	X
	Response	X	X	X	X	X	X	X	X
%Change	**Tumour**	X	X	X	X	X	X	X	X
	↑Response	↑ Breast p=0.065	↑ Breast p=0.048	↑ Breast p=0.228	X	X	↓ Breast p=0.140	↑ Breast p=0.093	X

Table 3-12: ^{18}F-Fluoride PET Per-lesion analysis; All Parameters; Summary Table

Although the clinical use of ^{18}F-Fluoride PET for assessment of bone metastases is widespread, there is no evidence for assessment other than qualitative analysis for these scans. Hiller et al. reported 59% of patients in their retrospective study had treatment decisions changed following addition information from the ^{18}F-Fluoride PET[190]; it is not clear what information was used to effect these changes.

At the time of the study design there was little evidence to show ^{18}F-Fluoride PET quantification from baseline imaging had validity for response prediction. Zukotynski et al.[191], demonstrated no utility for predicting treatment outcome from baseline volumetric or SUV parameters in 9 prostate cancer patients with bone metastases. A non-cancer study reviewing the use of ^{18}F-Fluoride to monitor bone healing following complex orthopaedic procedures suggested ^{18}F-Fluoride could predict non-response to the orthopaedic interventions (surgical procedures rather than medical therapy)[192].

To my knowledge there is no literature reporting the successful quantification of baseline ^{18}F-Fluoride PET scans for prediction of a treatment response of bone metastases; the limited studies published have indeed found no value. This study has also demonstrated no value from the baseline parameters for prediction of the clinical treatment response.

There is more literature describing the clinical application of quantifying the changes in ^{18}F-Fluoride parameters. Responses to therapy for non-malignant bone conditions has been reported (e.g. osteopenia patients had an increase in Ki following anabolic therapy[193], and a rat model demonstrated a negative correlation with Ki and positive correlation with SUV following inducement of osteoporosis[194]). A decrease in ^{18}F-Fluoride uptake with successful therapy (correlated with the changes in tumour markers and markers of bone turnover) was shown in prostate cancer metastases by Cook et al., following therapy with Alpharadin[140], hypothesising the utility of ^{18}F-Fluoride PET for monitoring therapy of bone metastases. Apolo et al. confirmed a significant correlation between the SUV changes of ^{18}F-Fluoride PET, repeated at 6 and 12 months after initiating therapy, with PSA changes[195]. Takalkar et al. reported a decrease in ^{18}F-

Fluoride SUV_{max} after 2 and 6 cycles of Alpharadin for bone metastases from breast cancer[196], suggesting a treatment response can be assessed at a clinically useful time.

A recent publication compared ^{18}F-Fluoride and ^{18}F-FDG PET imaging for 34 patients with multiple myeloma who were imaged before and after autologous stem cell transplantation. The authors report Ki, $SUV_{average}$ and SUV_{max} from ^{18}F-Fluoride PET all decreased following treatment. They report persistence of 81.5% of the baseline ^{18}F-Fluoride positive lesions despite 65% of these lesions becoming ^{18}F-FDG negative. The performance of ^{18}F-Fluoride PET was limited, and did not have accuracy for identifying response to therapy[197]. However, myeloma has a predominantly osteolytic action in bone, and although there will be osteoblast activity to some degree; bone scans and ^{18}F-Fluoride PET has limited utility for imaging myeloma.

Yu et al. reported a decrease in SUV_{max} 12 weeks after initiation of a tyrosine kinase inhibitor for prostate cancer bone metastases. A correlation between an increased SUV_{max} and a lower risk of PD was shown[198]; this may represent a flare response. Another group demonstrated a fall in SUV_{max} identified after 8-weeks and 24-weeks of Radium-223 for bone metastases from breast cancer in one patient with a treatment response[196].

In a study to compare the utility of ^{18}F-Fluoride and ^{18}F-FDG PET imaging for bone metastases from prostate cancer before and after therapy, Simoncic et al. reported a correlation between the changes in ^{18}F-Fluoride and ^{18}F-FDG[199]. There was no clinical correlation with treatment response. Interestingly the authors concluded up to an 80% discordance in spatial overlap of uptake from the two tracers; the stromal bone response resulting in ^{18}F-Fluoride uptake is not necessarily where the tumour cells are, indicated by the ^{18}F-FDG uptake[199].

In this study, an increase in SUV between the scans predicts for the treatment response for the breast cancer lesions (there were too few responses for analysis of the prostate lesions). $\%\Delta SUV_{max}$ and $\%\Delta SUV_{mean}$ have shown the most utility. A similar pattern in the data is seen for SUV_{peak}, but this does not approach statistical significance. This might suggest it is the SUV extremes (e.g. SUV_{max}, or the description of the whole SUV histogram with SUV_{mean}) with predictive clinical utility. Healing bone metastases are often associated with increased sclerosis from osteoblastic repair. 6 of the 14 (43%) patients with a clinical treatment response at 24-weeks had a measurable increase in SUV_{mean} between the scans; the literature suggests 6-23% of prostate cancer patients have a demonstrable flare response on a bone scan[200-203], and 12-29% of breast cancer patients with bone metastases[127,129].

The changes in SUV heterogeneity parameters also appear to suggest potential utility for predicting a treatment response in breast cancer lesions, with a decrease in $SUV_{entropy}$ and increase in SUV_{energy} suggesting a better outcome (i.e. a shift towards image homogeneity). This, however, is difficult to interpret. There is a strong and significant correlation between the %Δ SUV heterogeneity parameters and the %Δ SUV parameters. It is not clear what features of the image are contributing to the changes in the heterogeneity parameters, and they may simply be another way to measure a decrease in the range of SUV values across a VOI.

The changes in the volumetric parameters have demonstrated no utility. Although the %Δ SUV_{mean} predicted treatment response, this effect was not strong enough to overcome the lack of

association of the MTV with treatment response, and therefore TLA has not shown predictive benefit. ^{18}F-Fluoride PET uptake is an indirect measure of the tumour within bone; although a decrease in the number and extent of tumour cells following a treatment response might be expected, this may not translate into a change in the volume of bone influenced by the tumour and treatment activity. Indeed, a healing bone metastasis will still demonstrate osteoblastic activity for healing.

A fall in Ki has been demonstrated by other groups following successful therapy. Sachpekidis et al. measured a fall in Ki of bone lesions following therapy for myeloma[197]. In contrast, Yu et al. did not identify a difference in the Ki between tumour and normal bone following therapy[198]. They identified SUV_{max} as a better indicator of ^{18}F-Fluoride incorporation, but were not expecting this finding, attributing image noise to the limited precision of Ki estimates.

In conclusion, the per-lesion analysis of ^{18}F-Fluoride PET of breast cancer bone metastases has shown a 24-week treatment response can be identified at 8 weeks with an increase in SUV_{mean} and SUV_{max} between scans, probably representing a flare response of osteoblastic repair activity is being identified. Significant differences in the parameters between breast and prostate cancer have been identified.

Chapter 4 ^{18}F-Fluoride PET – Per-Lesion Analysis. Image quantification for survival analysis of patients with bone metastases from breast or prostate cancer

The methods of this chapter and the statistical approaches for this section can be reviewed in Chapter 2.3.3.

4.1 OS Analysis

62 lesions have been analysed (24 breast (no deaths), 38 prostate (15/38 in 3 patient deaths).

4.1.1 Tumour Group Analysis

The breast lesions have a significant OS benefit over the prostate lesions(Fig.4-1, KM analysis p=.00004).

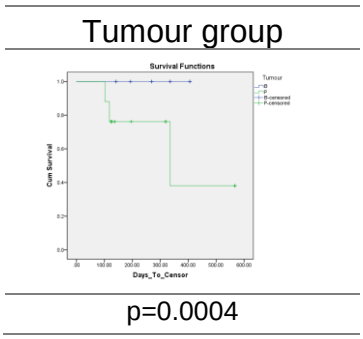


Figure 4-1: ^{18}F-Fluoride PET Per-lesion analysis; KM OS analysis comparing tumour groups (B=Breast cancer P=Prostate cancer) (p-values from Log Rank analysis)

4.1.2 Baseline Scan

4.1.2.1 SUV Parameters

	SUV$_{mean}$	**SUV$_{median}$**	**SUV$_{peak}$**
	Cox regression p-value [HR (95%CI)]		
n=62 15 events	**0.0002** [1.047 (1.022-1.073)]	**0.0002** [1.023 (1.011-1.035)]	**0.0002** [1.026 (1.012-1.04)]

Table 4-1: ^{18}F-Fluoride PET Per-lesion analysis; Baseline Scan; SUV Parameters; OS Analysis with univariate Cox regression analysis

A higher SUV is correlated with an increased risk of death (see Cox regression results in Tbl.4-1). KM OS analysis (Fig.4-2), show a significant OS advantage for lesions with a lower baseline SUV.

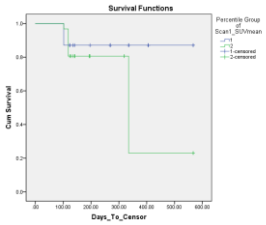
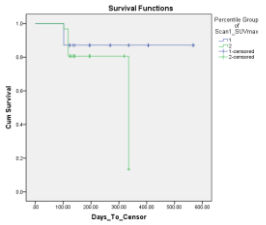
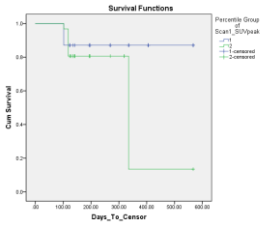
SUV_{mean}	SUV_{max}	SUV_{peak}
		
p=0.029 Low (1) - blue line High (2) - green line	p=0.018 Low (1) - blue line High (2) - green line	p=0.018 Low (1) - blue line High (2) - green line

Figure 4-2: ^{18}F-Fluoride PET Per-lesion analysis; Baseline Scan SUV Parameters; OS analysis with KM analysis and log rank for comparison; All lesions

Multivariate analysis is not feasible, but the prostate cancer lesions (higher SUV and poorer OS) can be interrogated separately (see Tbl.4-2 and Fig.4-3).

	SUV_{mean} **Prostate lesions only**	**SUV_{median}** **Prostate lesions only**	**SUV_{peak}** **Prostate lesions only**
	Cox regression p-value [HR (95%CI)]		
n=38, 15 events	**0.038** [1.03 (1.002-1.059)]	**0.029** [1.014 (1.001-1.027)]	**0.036** [1.016 (1.001-1.032)]

Table 4-2: ^{18}F-Fluoride PET Per-lesion analysis; Baseline Scan; SUV Parameters; OS Analysis with univariate Cox regression analysis; Prostate lesions only

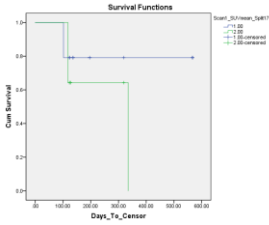
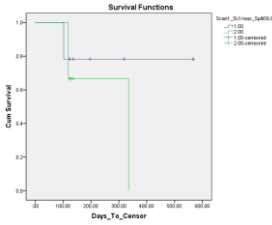
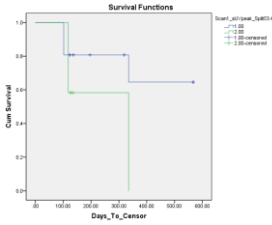
SUV_{mean} Prostate lesions only	SUV_{max} Prostate lesions only	SUV_{peak} Prostate lesions only
		
Dichotomised at 17.66 Low (1) - blue line High (2) - green line	Dichotomised at 56.8 Low (1) - blue line High (2) - green line	Dichotomised at 53.48 Low (1) - blue line High (2) - green line
p=0.048	p=0.074	p=0.036

Figure 4-3: ^{18}F-Fluoride PET Per-lesion analysis; Baseline Scan SUV Parameters; OS analysis with KM analysis and log rank for comparison; Prostate lesions only

For the prostate cancer lesions, a higher baseline SUV correlates significantly with a higher risk of death (results in Tbl.4-2), confirmed with KM analysis (Fig.4-3).

4.1.2.2 Volumetric Parameters

	MTV	TLA
	Cox regression p-value [HR (95%CI)]	
n=62 15 events	**0.841** [1.001 (0.992-1.01)]	**0.109** [1 (1-1)]

Table 4-3: ^{18}F-Fluoride PET Per-lesion analysis; Baseline Scan; Volumetric Parameters; OS Analysis with univariate Cox regression analysis

Cox regression analysis (Tbl.4-3) shows no correlation between the volumetric parameters and OS. KM OS analysis can be reviewed in Fig.4-4.

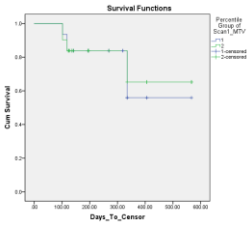
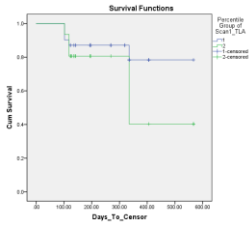
MTV	TLA
	
p=0.801	p=0.131
Low (1) - blue line High (2) - green line	Low (1) - blue line High (2) - green line

Figure 4-4: ^{18}F-Fluoride PET Per-lesion analysis; Baseline Scan Volumetric Parameters; OS analysis with KM analysis and log rank for comparison; All lesions

There is no significant separation of the survival curves for either parameter. Analysis of the prostate cancer lesions separately does not indicate prognostic benefit from the baseline volumetric parameters (Cox regression analysis HR 1.0 p=0.696).

4.1.2.3 Heterogeneity Parameters

	$SUV_{entropy}$	SUV_{energy}
	Cox regression p-value [HR (95%CI)]	
n=62, 15 events	**0.001** [3.062 (1.544-6.074)]	**0.01** [4e-9 (0-0.011)]

Table 4-4: ^{18}F-Fluoride PET Per-lesion analysis; Baseline Scan; SUV Heterogeneity Parameters; OS Analysis with univariate Cox regression analysis

With Cox regression analysis (Tbl.4-4) a higher $SUV_{entropy}$ correlates with a higher risk of heath (p=0.001); a higher SUV_{energy} correlates with a lower risk of death (p=0.01), suggesting an OS benefit for patients with a less random and more uniform SUV distribution. These findings are corroborated by KM OS analysis (Fig.4-5). However, subgroup analysis of the prostate cancer patients demonstrates no association. The baseline heterogeneity parameters demonstrated no significant difference in the distribution of the SUV heterogeneity parameters between the tumour types, and analysis of the prostate cancer lesions alone may be impaired by a reduction of the sample group; multivariate analysis is not feasible with so few data to analyse this further.

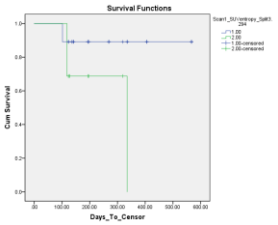
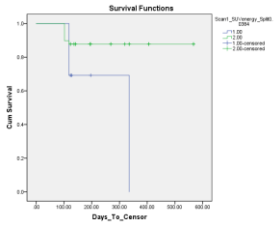
$SUV_{entropy}$	SUV_{energy}
	
Dichotomised at 3.294 Low (1) - blue line High (2) - green line	Dichotomised at 0.0384 Low (1) - blue line High (2) - green line
p=0.0002	p=0.0004

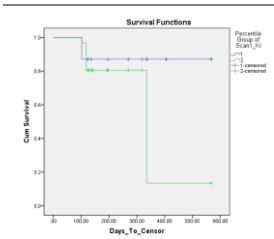
Figure 4-5: ^{18}F-Fluoride PET Per-lesion analysis; Baseline Scan SUV Heterogeneity Parameters; OS analysis with KM analysis and log rank for comparison; All lesions

4.1.2.4 Ki

	Ki
	Cox regression p-value [HR (95%CI)]
n=62, 15 events	**0.001** [21.29 (3.802-119.2)]

Table 4-5: ^{18}F-Fluoride PET Per-lesion analysis; Baseline Scan; Ki; OS Analysis with univariate Cox regression analysis

A higher Ki is correlated with an increased risk of death (Tbl.4-5 - HR 21.29, p=0.001). The data has been dichotomised at the median, and KM analysis (Fig.4-6)) demonstrates a statistically significant separation of the survival curves (p=0.018) with a survival advantage for patients with lesions with a lower baseline PET Ki.



p=0.018

Figure 4-6: ^{18}F-Fluoride PET Per-lesion analysis; Baseline Scan Ki; OS analysis with KM analysis and log rank for comparison; All lesions

Cox regression analysis of only the prostate cancer patients identifies consistency of this finding (HR 8.276 (1.332-51.414 95%CI), p=0.023, indicating a higher Ki is associated with a significantly increased risk of death for the prostate patients.

4.1.3 %Δ Between Scans

4.1.3.1 %ΔSUV Parameters

	%ΔSUV$_{mean}$	%ΔSUV$_{max}$	%ΔSUV$_{peak}$
	Cox regression p-value **[HR (95%CI)]**		
n=53 (B=22, P=31) 10 events	**0.104** [0.982 (0.96-1.004)]	**0.007** [0.962 (0.936-0.989)]	**0.008** [0.964 (0.939-0.99)]

Table 4-6: ^{18}F-Fluoride PET Per-lesion analysis; %Δ SUV Parameters; OS Analysis with univariate Cox regression analysis

For these lesions, there is no correlation between the size of change of SUV$_{mean}$ between the scans and the risk of death (Tbl.4-6). However, %ΔSUV$_{max}$ and %ΔSUV$_{peak}$ appear to be predictive parameters; an increase in the SUV parameters between scans is associated with a lower risk of death. This may represent a flare response. KM analysis has been performed using the data (Fig.4-7).

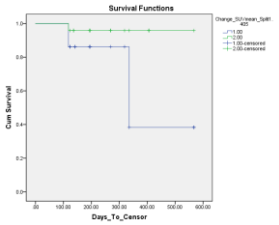
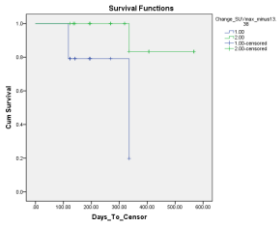
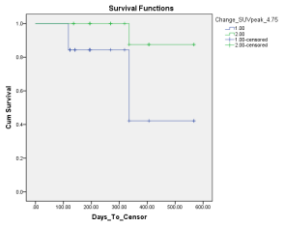
%ΔSUV$_{mean}$	%ΔSUV$_{max}$	%ΔSUV$_{peak}$
		
Dichotomised at 1.405% Low (1) - blue line High (2) - green line	Dichotomised at -13.38% Low (1) - blue line High (2) - green line	Dichotomised at 4.75% Low (1) - blue line High (2) - green line
p=0.016	p=0.001	p=0.014

Figure 4-7: ^{18}F-Fluoride PET Per-lesion analysis; %Δ SUV Parameters; OS analysis with KM analysis and log rank for comparison; All lesions

Increases greater than 1.405% in SUV$_{mean}$ (p=0.016), increases in SUV$_{max}$ (or decreases smaller than 13.28%) (p=0.001), and SUV$_{peak}$ increases above 4.75% (p=0.014) are all associated with better OS.

However, the breast cancer patients on average had a larger increase in the SUV parameters between the scans. KM analysis of just the prostate cancer lesions can be reviewed in Fig.4-8, demonstrating persistence of the OS benefit for lesions with an increase in SUV between the scans. The findings are statistically significant for %Δ SUV$_{max}$ and %Δ SUV$_{peak}$ (p=0.002 and 0.014, respectively), although the survival curves do meet, and with so few events this would require a larger sample for ratification. This probably represents healing osteoblastic activity secondary to a treatment response.

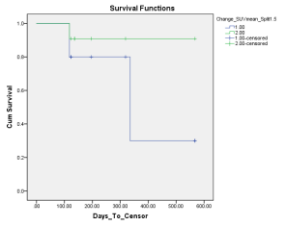
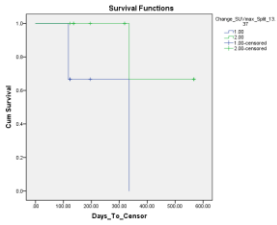
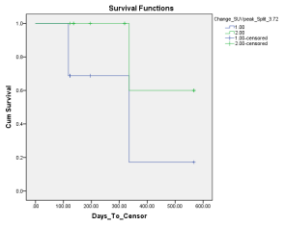
%ΔSUV$_{mean}$ Prostate lesions only	%ΔSUV$_{max}$ Prostate lesions only	%ΔSUV$_{peak}$ Prostate lesions only
		
Dichotomised at +1.5% Low (1) - blue line High (2) - green line	Dichotomised at -13.37% Low (1) - blue line High (2) - green line	Dichotomised at -3.72% Low (1) - blue line High (2) - green line
p=0.232	p=0.002	p=0.014

Figure 4-8: ^{18}F-Fluoride PET Per-lesion analysis; %Δ SUV Parameters; OS analysis with KM analysis and log rank for comparison; Prostate lesions only

4.1.3.2 %Δ Volumetric Parameters

	%ΔMTV	%ΔTLA
	Cox regression p-value [HR (95%CI)]	
n=53 (B=22, P=31) **10 events**	**0.315** [1.003 (0.997-1.008)]	**0.961** [1 (0.99-1.01)]

Table 4-7: ^{18}F-Fluoride PET Per-lesion analysis; %Δ Volumetric Parameters; OS Analysis with univariate Cox regression analysis

Cox regression (Tbl.4-7) does not show any correlation between %Δ volumetric parameters and the risk of death. KM analysis has been performed (Fig.4-9), but no significant separation of the survival curves is identified. The volumetric parameters have no predictive benefit in these lesions for assessing patient OS. Analysis of just the prostate cancer lesions only also identifies no utility (%Δ MTV p=0.396 log rank analysis, %Δ TLA p=0.358).

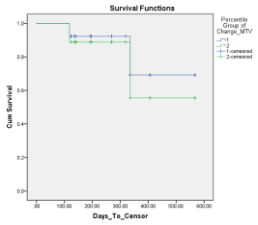
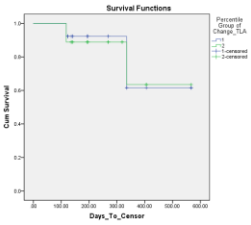
%ΔMTV	%ΔTLA
	
Low (1) - blue line High (2) - green line	Low (1) - blue line High (2) - green line
p=0.508	p=0.853

Figure 4-9: ^{18}F-Fluoride PET Per-lesion analysis; %Δ Volumetric Parameters; OS analysis with KM analysis and log rank for comparison; All lesions

4.1.3.3 %Δ Heterogeneity Parameters

	%ΔSUV$_{entropy}$	**%ΔSUV$_{energy}$**
	Cox regression p-value [HR (95%CI)]	
n=53 (B=22, P=31) 10 events	0.09 [0.957 (0.909-1.007)]	0.024 [1.016 (1.002-1.03)]

Table 4-8: ^{18}F-Fluoride PET Per-lesion analysis; %Δ SUV Heterogeneity Parameters; OS Analysis with univariate Cox regression analysis

Cox regression analysis (Tbl.4-8) suggests a correlation between the %ΔSUV$_{entropy}$ and OS, although the hazard-ratio confidence interval and does not quite achieve statistical significance (p=0.09). The trend suggests a higher %ΔSUV$_{entropy}$ and lower %Δ$_{energy}$ is associated a decreased risk of death. This may be identifying a flare response in responding lesions causing more heterogeneity in the SUV uptake across the lesions. KM OS analysis can be reviewed below (Fig.4-10):

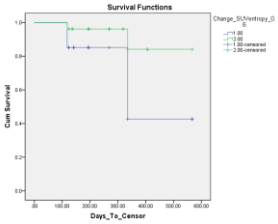
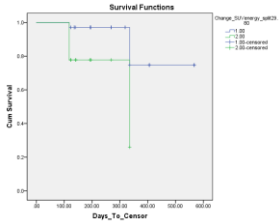
%ΔSUV$_{entropy}$	%ΔSUV$_{energy}$
	
Dichotomised at +0.6% Low (1) - blue line High (2) - green line	*Dichotomised at +29.8%* Low (1) - blue line High (2) - green line
p=0.04	p=0.008

Figure 4-10: ^{18}F-Fluoride PET Per-lesion analysis; %Δ SUV Heterogeneity Parameters; OS analysis with KM analysis and log rank for comparison; All lesions

A %Δ SUV$_{entropy}$ >0.6% (p=0.04) and %ΔSUV$_{energy}$ ≤ 29.8% (p=0.008) is associated with an OS benefit (p=0.04). The breast cancer lesions had a larger increase in the SUV heterogeneity parameters between the scans, although this not statistically significant. Analysis of the prostate cancer patients alone reveals an OS advantage for patents with a lower %ΔSUV$_{energy}$ (Fig.4-11). However, the curves meet, and with so few events conclusions cannot be confidently drawn. Further analysis with sufficient data for multivariate analysis is necessary.

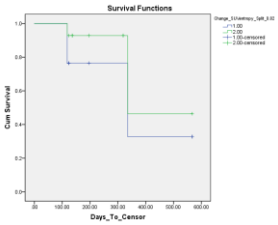
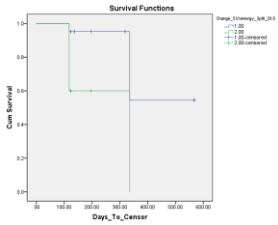
%ΔSUV$_{entropy}$ Prostate lesions only	%ΔSUV$_{energy}$ Prostate lesions only
	
Dichotomised at +0.82% Low (1) - blue line High (2) - green line	Dichotomised at +29.8% Low (1) - blue line High (2) - green line
p=0.264	p=0.005

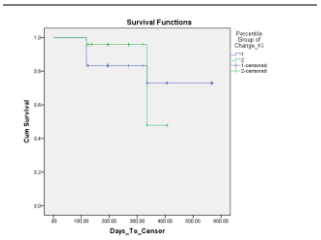
Figure 4-11: ^{18}F-Fluoride PET Per-lesion analysis; %Δ SUV Parameters; OS analysis with KM analysis and log rank for comparison; Prostate lesions only

4.1.3.4 %Δ Ki

	%ΔKi
	Cox regression p-value [HR (95%CI)]
n=48, 10 events	**0.205** [0.986 (0.965-1.008)]

Table 4-9: ^{18}F-Fluoride PET Per-lesion analysis; %Δ Ki; OS Analysis with univariate Cox regression analysis

There is no demonstrable statistically significant correlation between the size of change of the Ki parameter and the risk of death (Tbl.4-9). KM analysis has also not demonstrated statistically significant separation of the survival curves (Fig.4-12). There is also no evidence of OS prediction utility when only the prostate cancer lesions are analysed (p=0.880). The change in Ki has not been a valuable OS predictor for these lesions



Low (1) - blue line
High (2) - green line

p=1.0

Figure 4-12: ^{18}F-Fluoride PET Per-lesion analysis; %Δ Ki; OS analysis with KM analysis and log rank for comparison; All lesions

4.2 PFS Analysis

4.2.1 Tumour Group Analysis

60 lesions (24 breast (10/24 in 2 patients with PD), 38 prostate cancer lesions (35/38 in 6 patients with PD) have been analysed. KM PFS analysis based on the tumour group (Fig.4-13; p=0.0004) confirms a significant PFS benefit for the breast cancer lesions.

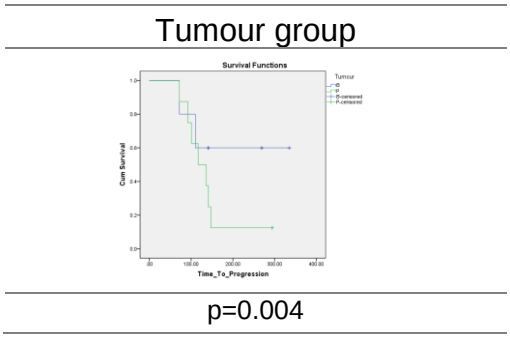


Figure 4-13: ^{18}F-Fluoride PET Per-lesion analysis; KM PFS analysis comparing tumour groups (B=Breast cancer P=Prostate cancer) (p-values from Log Rank analysis)

4.2.2 Baseline Scan

4.2.2.1 SUV Parameters

	SUV$_{mean}$	**SUV$_{median}$**	**SUV$_{peak}$**
	Cox regression p-value [HR (95%CI)]		
n=60, 44 events	**0.189** [1.015 (0.993-1.037)]	**0.132** [1.007 (0.998-1.015)]	**0.151** [1.008 (0.997-1.018)]

Table 4-10: ^{18}F-Fluoride PET Per-lesion analysis; Baseline Scan SUV Parameters; PFS Analysis with univariate Cox regression analysis

Cox regression analysis (Tbl.4-10) shows no correlation between the magnitude of the baseline SUV parameters and the PFS. KM analysis can be reviewed below (Tbl.4-11). There is a significant separation when using SUV$_{max}$ (p=0.034) and SUV$_{peak}$ (p=0.022), a PFS benefit for those lower SUV. However, KM analysis of the prostate cancer lesions only does not reveal PFS predictivity utility using the baseline SUV parameters ((SUV$_{mean}$ p=0.868, SUV$_{max}$ p=0.511, SUV$_{peak}$ p=0.761), nor for the breast cancer lesions (SUV$_{mean}$ p=0.248, SUV$_{max}$ p=0.440, SUV$_{peak}$ p=0.614). The SUV parameters are not independent predictors of PFS for these lesions.

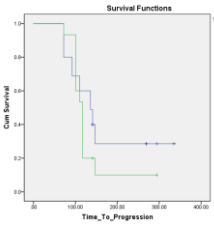
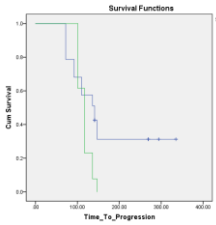
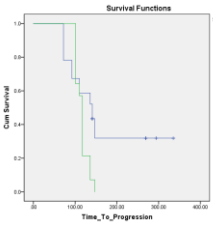
	SUV_{mean}	SUV_{median}	SUV_{peak}
	Median PFS in days (95% CI)		
KM			
Cut-off	16.57	56.96	40.47
(1) ≤	136 (113.4-158.6)	141 (111.6-170.4)	141 (111.8-170.2)
(2) >	117 (108.9-125.1)	117 (107.5-126.5)	117 (112.8-121.2)
p-Value (Log Rank)	0.141	0.034	0.022

Table 4-11: ^{18}F-Fluoride PET Per-lesion analysis; Baseline SUV Parameters; PFS analysis with KM analysis and log rank for comparison; All lesions

4.2.2.2 Volumetric Parameters

	MTV	**TLA**
	Cox regression p-value [HR (95%CI)]	
n=60, 44 events	**0.541** [1.002 (0.996-1.007)]	**0.36** [1 (1-1)]

Table 4-12: ^{18}F-Fluoride PET Per-lesion analysis; Baseline Scan Volumetric Parameters; PFS Analysis with univariate Cox regression analysis

Cox regression analysis (Tbl.4-12) does not identify a correlation between the size of the volumetric parameters and the risk of PD. The data has been dichotomised for KM analysis (Tbl.4-13).

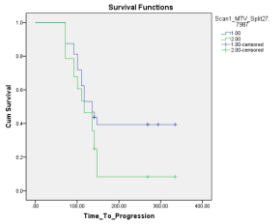
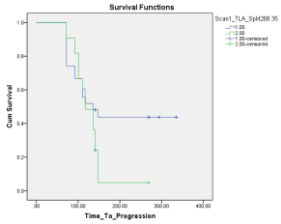
	MTV	**TLA**
	Median PFS in days (95% CI)	
KM		
Cut-off	27.80	268.35
(1) ≤	136 (107.4-164.6)	136 (76.8-195.2)
(2) >	117 (91.1-142.9)	117 (98.7-135.3)
p-Value (Log Rank)	0.045	0.05

Table 4-13: ^{18}F-Fluoride PET Per-lesion analysis; Baseline Volumetric Parameters; PFS analysis with KM analysis and log rank for comparison; All lesions

A small but significant 19-day benefit in PFS for lesions smaller than 27.8cm^3 (p=0.045) and for lesions with a TLA ≤ 268.35 (0.05) has been identified. The prostate cancer lesions, on average, had higher MTV and TLA parameters than the breast cancer lesions. Analysis of the breast cancer lesions suggests a trend for a PFS benefit (median 19-day) for the smaller breast cancer lesions (p=0.093), and the lesions with a lower calculated TLA (p=0.126) – see Tbl.4-14.

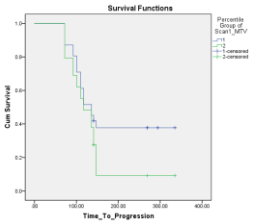
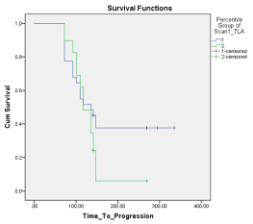
	MTV Breast only	TLA Breast only
	Median PFS in days (95% CI)	
KM Plots		
Median (cut-off)	23.62cm^3	210.55
(1) ≤	136 (107.9-164.1)	136 (105.9-166.1)
(2) >	117 (89.6-144.4)	117 (99.9-134.1)
p-Value (Log Rank)	0.093	0.126

Table 4-14: ^{18}F-Fluoride PET Per-lesion analysis; Baseline SUV Parameters; PFS analysis with KM analysis and log rank for comparison; Breast lesions only

4.2.2.3 Heterogeneity Parameters

	$SUV_{entropy}$	SUV_{energy}
	Cox regression p-value [HR (95%CI)]	
n=60, 44 events	**0.294** [1.281 (0.807-2.035)]	**0.563** [0.121 (0-155.445)]

Table 4-15: ^{18}F-Fluoride PET Per-lesion analysis; Baseline Scan SUV Heterogeneity Parameters; PFS Analysis with univariate Cox regression analysis

Cox regression analysis (Tbl.4-15) does not demonstrate any significant correlation for these lesions. The data has been dichotomised at the median for KM PFS analysis (Tbl.4-16)), but no significant PFS prognostic utility has been identified. Analysis of the breast cancer lesions independently does not demonstrate significant separation of the KM PFS curves.

There is no demonstrable PFS predictive benefit for these lesions using the baseline SUV heterogeneity parameters.

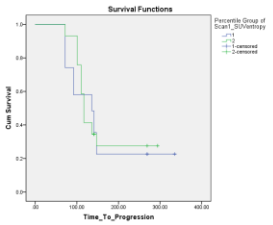
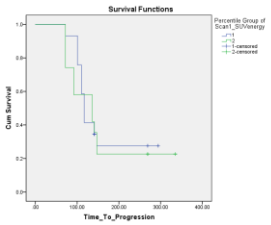
	$SUV_{entropy}$	SUV_{energy}
	Median PFS in days (95% CI)	
KM Plots		
Median (cut-off)	2.75	0.0761
(1) ≤	136 (97.8-174.2)	117 (109.7-124.3)
(2) >	117 (109.7-124.3)	136 (88.6-183.4)
p-Value (Log Rank)	0.773	0.695

Table 4-16: ^{18}F-Fluoride PET Per-lesion analysis; Baseline SUV Heterogeneity Parameters; PFS analysis with KM analysis and log rank for comparison; All lesions

4.2.2.4 Ki

	Ki
	Cox regression p-value [HR (95%CI)]
n=60, 44 events	**0.302** [2.249 (0.483-10.461)]

Table 4-17: ^{18}F-Fluoride PET Per-lesion analysis; Baseline Scan Ki; PFS Analysis with univariate Cox regression analysis

There is no correlation identified with Cox regression analysis (Tbl.4-17). KM analysis (Tbl.4-18)) has also not demonstrated an association between Ki and PFS.

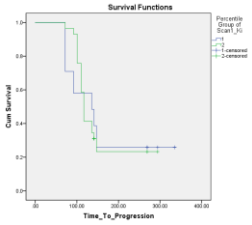
	Ki
	Median PFS in days (95% CI)
KM Plots	
Median (cut-off)	0.16
(1) ≤	136 (91.5-180.5)
(2) >	117 (109.7-124.3)
p-Value (Log Rank)	0.911

Table 4-18: ^{18}F-Fluoride PET Per-lesion analysis; Baseline Ki; PFS analysis with KM analysis and log rank for comparison; All lesions

Analysis of the prostate cancer and breast cancer lesions independently also does not demonstrate a significant separation of the PFS curves (breast cancer, p=0.911).).

4.2.3 %Δ Parameters Between Scans

53 lesions (22 breast (10/22 in 2/5 patients with PD), 31 prostate cancer lesions (29/31 in 6/7 patients with PD) have been analysed.

4.2.4 %ΔSUV Parameters

	%ΔSUV$_{mean}$	**%ΔSUV$_{median}$**	**%ΔSUV$_{peak}$**
	Cox regression p-value [HR (95%CI)]		
n=53, 39 events	**0.444** [1.003 (0.996-1.009)]	**0.201** [1.005 (0.997-1.014)]	**0.545** [1.003 (0.994-1.012)]

Table 4-19: ^{18}F-Fluoride PET Per-lesion analysis; %Δ SUV Parameters; PFS Analysis with univariate Cox regression analysis

For these lesions, no correlation between the size of change in the SUV parameters between scans and the risk of PD has been identified (Tbl.4-19).

KM analysis has been performed for all the lesions together, revealing no significant separation of the dichotomised PFS survival curves (Tbl.4-20).

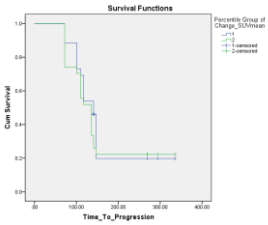
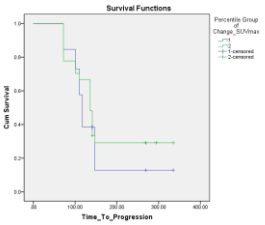
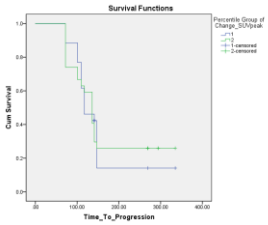
	%ΔSUV$_{mean}$	**%ΔSUV$_{median}$**	**%ΔSUV$_{peak}$**
	Median PFS in days (95% CI)		
KM Plots			
Median	0.92%	-11.07%	-7.67%
(1) ≤	141 (126.7-155.3)	117 (110.2-123.8)	117 (86.1-147.9)
(2) >	136 (115.2-156.8)	136 (118.5-153.5)	136 (117-155)
p-Value	0.432	0.461	0.998

Table 4-20: ^{18}F-Fluoride PET Per-lesion analysis; %Δ SUV parameters; PFS analysis with KM analysis and log rank for comparison; All lesions

There is no PFS benefit identified when only the prostate cancer lesions are analysed (%ΔSUV$_{mean}$ p=0.665, %ΔSUV$_{max}$ p=0.243, %ΔSUV$_{peak}$ p=0.397), or the breast cancer lesions ((%ΔSUV$_{mean}$ p=0.298, %ΔSUV$_{max}$ p=0.204, %ΔSUV$_{peak}$ p=0.204).

4.2.4.1 %Δ Volumetric Parameters

For these lesions, the percentage change in the volumetric parameters between scans has no demonstrable role in predicting for PFS. The results of Cox regression analyses (Tbl.4-21) and KM analyses (Tbl.4-22) can be reviewed below.

	%ΔMTV	%ΔTLA
	Cox regression p-value [HR (95%CI)]	
n=53, 39 events	**0.53** [0.999 (0.994-1.003)]	**0.514** [1.002 (0.996-1.008)]

Table 4-21: ^{18}F-Fluoride PET Per-lesion analysis; %Δ Volumetric Parameters; PFS Analysis with univariate Cox regression analysis

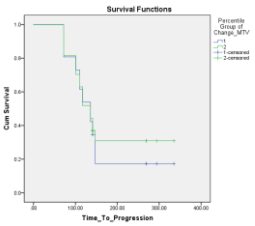
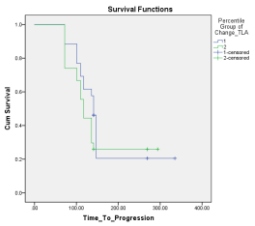
	%ΔMTV	%ΔTLA
	Median PFS in days (95% CI)	
KM Plots		
Median (cut-off)	1.98%	15.04%
(1) ≤	136 (110.3-161.7)	141 (135.3-146.7)
(2) >	136 (109.7-162.3)	117 (105.2-128.8)
p-Value (Log Rank)	0.610	0.355

Table 4-22: ^{18}F-Fluoride PET Per-lesion analysis; %Δ Volumetric parameters; PFS analysis with KM analysis and log rank for comparison; All lesions

There is no demonstrable predictive utility for these lesions using the %Δ volumetric parameters of ^{18}F-Fluoride PET. There is also no utility identified when only the prostate cancer lesions were analysed (%ΔMTV p=0.175, %ΔTLA p=0.409). The analysis (Tbl.4-23) of just the breast cancer lesions shows a trend towards significance for a PFS benefit for the lesions with a more negative (≤+21.91%) change in TLA between scans (p=0.078)

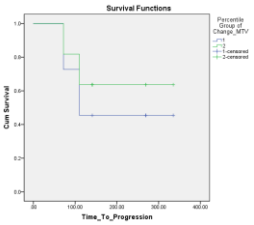
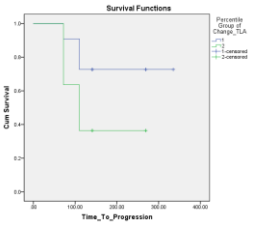
	%ΔMTV (breast only)	%ΔTLA (breast only)
KM		
Key	Low (1) - blue line High (2) - green line	
Median	+9.18%%	+21.91%%
p-value (log rank)	0.411	0.078

Table 4-23: ^{18}F-Fluoride PET Per-lesion analysis; %Δ Volumetric parameters; PFS analysis with KM analysis and log rank for comparison; Breast lesions only

4.2.4.2 %Δ Heterogeneity Parameters

The %ΔSUV heterogeneity parameters have no demonstrable role in predicting for PFS. The results of Cox regression analyses (Tbl.4-24) and KM analyses (Tbl.4-25) can be reviewed below.

	%ΔSUV$_{entropy}$	**%ΔSUV$_{energy}$**
	Cox regression p-value [HR (95%CI)]	
n=53, 39 events	**0.462** [1.007 (0.989-1.025)]	**0.334** [0.997 (0.99-1.003)]

Table 4-24: ^{18}F-Fluoride PET Per-lesion analysis; %Δ SUV Heterogeneity Parameters; PFS Analysis with univariate Cox regression analysis

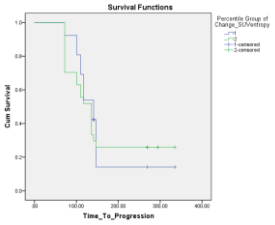
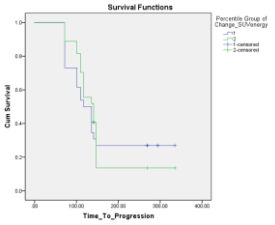
	%ΔSUV$_{entropy}$	**%ΔSUV$_{energy}$**
	Median PFS in days (95% CI)	
KM PFS Plots		
Median (cut-off)	0.26%	3.84%
(1) ≤	141 (119.1-162.9)	117 (92-142)
(2) >	136 (115.2-156.8)	141 (111-171)
p-Value (Log Rank)	0.657	0.747

Table 4-25: ^{18}F-Fluoride PET Per-lesion analysis; %Δ SUV Heterogeneity parameters; PFS analysis with KM analysis and log rank for comparison; All lesions

KM analysis of the breast cancer patients also shows no statistically significant association between the %Δ SUV heterogeneity parameters and PFS, but there is perhaps a trend towards a PFS benefit for lesions with a decrease in SUV$_{entropy}$ and increase in SUV$_{energy}$ between the scans (Tbl.4-26), i.e. a shift towards voxel homogeneity of distribution.

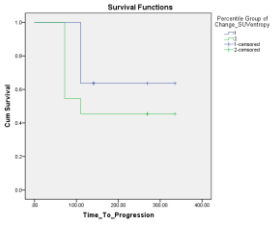
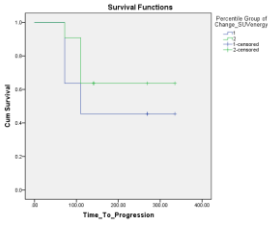
	%ΔSUV$_{entropy}$ **Breast only**	**%ΔSUV$_{energy}$** **Breast only**
	Median PFS in days (95% CI)	
KM PFS plots		
Key	Low (1) - blue line High (2) - green line	
Median (cut-off)	+1.76%	+1.92%
p-Value (Log Rank)	0.204	0.298

Table 4-26: ^{18}F-Fluoride PET Per-lesion analysis; %Δ SUV Heterogeneity parameters; PFS analysis with KM analysis and log rank for comparison; Breast lesions only

4.2.4.3 %ΔKi

For these lesions, the %ΔKi has no demonstrable role in predicting PFS. The results of Cox regression analysis (Tbl.4-27) and KM analysis (Tbl.4-28) can be reviewed below.

	%ΔKi
	Cox regression p-value [HR (95%CI)]
n=48, 39 events	**0.508** [0.997 (0.99-1.005)]

Table 4-27: ^{18}F-Fluoride PET Per-lesion analysis; %Δ Ki; PFS Analysis with univariate Cox regression analysis

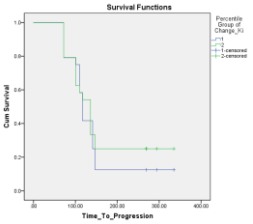
	%ΔKi
	Median PFS in days (95% CI)
KM Plot	
Median (cut-off)	0.15%
(1) ≤	117 (108.7-125.3)
(2) >	136 (116.4-155.6)
p-Value (Log Rank)	0.539

Table 4-28: ^{18}F-Fluoride PET Per-lesion analysis; %Δ Ki; PFS analysis with KM analysis and log rank for comparison; All lesions

There is no demonstrable predictive utility identified either with %ΔKi when only the prostate cancer lesions are analysed (median dichotomisation for KM analysis, p=0.361), or breast cancer lesions (p=0.383).

4.3 Discussion

	SUV_{mean}	SUV_{max}	SUV_{peak}	MTV	TLA	$SUV_{entropy}$	SUV_{energy}	Ki
↑OS	↓ All pts HR 1.047 p=0.0002 PCa – HR1.03, p=0.038	↓ All pts HR 1.023 p=0.0002 PCa – HR1.014, p=0.029	↓ All pts HR 1.026p=0.0002 PCa – HR1.016, p=0.036	X	X	↓ p=0.0002	↑ p=0.0004	↓ All pts – p=0.001 PCa – p=0.023
↑PFS	X	X	X	PCa – X BCa - ↓ p=0.093	PCa – X BCa - ↓ p=0.126	X	X	X

Table 4-29: ^{18}F-Fluoride PET Per-lesion analysis; Survival Analyses; Baseline Scan; Summary Table

	$\%\Delta SUV_{mean}$	$\%\Delta SUV_{max}$	$\%\Delta SUV_{peak}$	%ΔMTV	%ΔTLA	$\%\Delta SUV_{entropy}$	$\%\Delta SUV_{energy}$	%ΔKi
↑OS	↑ All – p=0.016 PCa – p=0.048	↑ All – p=0.001 PCa – p=0.074	↑ All – p=0.014 PCa – p=0.036	X	X	↑ All – p=0.041 PCa – p=0.364	↓ All – p=0.008 PCa – p=0.005	X
↑PFS	X	X	X	X	↓ BCa p=0.078	↓ Bca p=0.204	↑ BCa p=0.298	X

Table 4-30: ^{18}F-Fluoride PET Per-lesion analysis; Survival Analyses; %Δ Parameters; Summary Table

This chapter has suggested prognostic utility from ^{18}F-Fluoride PET scans of bone metastases, both from the pre-treatment scans, and with %Δ parameters. There is limited corroborative evidence in the literature to compare these findings.

Yu et al. published in 2015 a study of 17 patients (12 with a second ^{18}F-Fluoride PET scan at 12 weeks following treatment) with bone metastases from prostate cancer, due to start treatment with a tyrosine kinase inhibitor. The baseline SUV_{max} and Ki did not correlate with the PFS of these patients, but an increase in SUV_{max} between scans was associated with a lower risk of PD (HR 0.91, p=0.056)[198].

In the same year, another study of prostate cancer bone metastases was published by Etchebehere et al. They performed an ^{18}F-Fluoride PET before treatment with the radionuclide treatment Alpharadin. They concluded no prognostic utility of SUV_{mean} and SUV_{max}, but the measures of skeletal burden were independent OS predictors (HR 5.99, p=0.02)[204]. Prognostic utility of volumetric parameters was also shown by Lindgren Belal et al. in 2017; a study of 48 prostate cancer patients showed baseline measurements of the full skeletal burden of bone metastases predicted an OS benefit for patients with less than 39% of skeletal involved involvement[205].

The lack of prognostic utility of ^{18}F-Fluoride quantification parameters before therapy has also been suggested by Piccardo et al. They compared the role of baseline ^{18}F-Fluoride and ^{18}F-FDG PET imaging of 32 breast cancer patients with bone metastases; despite the higher diagnostic sensitivity of the ^{18}F-Fluoride imaging, none of the ^{18}F-Fluoride quantification parameters were associated with PFS or OS[206].

SUV quantification parameters have been shown to have possible utility. Apolo et al. reported a prospective pilot study of 30 patients with prostate cancer, imaging them with ^{18}F-Fluoride PET before treatment, at 6 months and again at 12 months. The patients were treated with a wide range of therapies. Baseline SUV_{max} was not a significant predictor of OS, but the maximal change at 6 months correlated with survival (HR 1.23, p=0.017). The change in SUV of the bone metastases correlated with the change in patients' PSA at 6 and 12 months, and an increase in SUV greater than 57% at 6 months correlated with PD. A significant increase in SUV between 6 and 12 months was seen in those patients who had PD at 12 months[195].

This thesis seems unique in identifying prognostic utility from semi-quantitative parameters from baseline ^{18}F-Fluoride PET of bone. 31 prostate cancer lesions have been imaged before treatment and again after 8 weeks of therapy. For the prostate cancer lesions a higher $SUV_{mean/max/peak}$ at baseline was correlated with OS (SUV_{mean} HR 1.03, p=0.038, SUV_{max} HR 1.014 p=0.029). These hazard ratios are small, but statistically significant, suggesting a signal with clinical relevance can be quantified. The baseline imaging also demonstrated OS prediction from a decrease in $SUV_{entropy}$ (p=0.0002) (towards image homogeneity), rise in SUV_{energy} (0.0004) (towards image homogeneity) and an increase in Ki (p=0.023). However, there are strong and significant correlations between all these parameters (negative correlation for SUV_{energy}) and the SUV parameters. There are too few lesions and events to warrant a multivariate analysis, and therefore it is unclear which of these parameters yields more prognostic accuracy.

24 breast cancer lesions were imaged with baseline imaging, and 22 had data available for calculation of the %Δ parameters. The baseline parameters suggested no utility of the SUV parameters for PFS prognostication, but the smaller lesions, measured with MTV, had a trend towards a PFS benefit (p=0.093). The whole skeletal burden of bone metastases has been shown to have prognostic relevance[204,205]. It is not clear if the size of individual lesions correlates with the skeletal burden. Up to 5 lesions per patient were identified; the largest and most avid lesions were selected. The total skeletal burden of disease in these patients could be assessed retrospectively to assess whether the individual size of the lesions is an independent predictor of PFS. The TLA parameter also shows a possible association with PFS, but TLA is closely correlated with MTV, being the product of MTV and SUV_{mean}, and tumour volume probably has the dominant prognostic effect for these lesions.

The %Δ parameters do show an association with OS of the prostate cancer lesions. An increase in SUV mean/max/peak suggests an OS benefit (p=0.048, 0.074, 0.036 respectively); this may represent a flare osteoblastic response. An increase in $SUV_{entropy}$ and decrease in SUV_{energy} (towards image homogeneity) also suggest OS prognostication, but there is a strong and significant correlation between these parameters and the SUV parameters (negative correlation for %Δ SUV_{energy}), and it is not possible with so few patients and lesions to perform multivariate analysis.

The change in SUV parameters between scans also have not predicted PFS. The change in MTV has not shown utility, but the TLA suggests a fall in TLA between scans predicts PFS of the breast cancer lesions (p=0.078). Response analysis (See Ch.3) showed an increase in SUV

between scans suggested a treatment response, and there was no demonstrable difference between the MTV of those with a treatment response and those with no response to treatment. Therefore, a fall in TLA being associated with PFS benefit was unexpected. This may be a statistical anomaly; the lesions were identified in only 5 patients, and the PFS analysis of the lesions will be unduly influenced by the prolonged PFS of even one patient. None of the %Δ parameters are associated with PFS in this analysis.

This study has identified novel prognostic information from ^{18}F-Fluoride PET quantification of bone metastases. Prostate cancer metastases with a higher baseline SUV, or with heterogeneity parameters measuring more homogeneity, have a significant OS benefit . An increase in SUV by 8-weeks of the prostate cancer lesions has shown a significant OS benefit, corroborating the literature[195].

Chapter 5 ^{18}F-Fluoride PET Per-Patient Analysis: Impact of VOI segmentation on image quantification of bone metastases

The methods of this chapter and the statistical approaches for this section can be reviewed in Chapter 2.3.4.

5.1 Baseline Scan

14 patients had baseline ^{18}F-Fluoride PET imaging, 5 with breast cancer, 9 with prostate cancer.

5.1.1 SUV Parameters

	SUV_{mean}	SUV_{max}	SUV_{peak}
		Mean (Min-Max) [Median; SD]	
VOI_{Manual}	**8.98** (3.91-23.4) [6.57; 5.81]	**45.23** (17.22-130.63) [34.39; 32.06]	**34.59** (11.68-101.48) [27.12; 25.06]
VOI_{27}	**16.07** (6.68-43.9) [10.13; 11.45]	**44.92** (17.22-130.63) [34.39; 31.67]	**34.32** (11.66-100.94) [27.19; 24.65]
VOI_{40}	**18.96** (8.06-51.01) [12.47; 13.13]	**44.92** (17.22-130.63) [34.39; 31.67]	**34.24** (11.84-101.24) [27.06; 24.75]
VOI_{50}	**21.08** (8.96-56.65) [14.27; 14.41]	**44.92** (17.22-130.63) [34.39; 31.67]	**34.9** (11.95-104.99) [27.03; 25.66]
VOI_{FLAB}	**15.05** (5.27-43.18) [11.27; 11.12]	**44.96** (17.22-130.63) [34.39; 31.64]	**34.25** (11.66-101.27) [27.07; 24.7]

Table 5-1: ^{18}F-Fluoride PET; Per-Patient Analysis; Baseline Scan; SUV Parameters; Descriptive statistics for all VOI methods and all patients

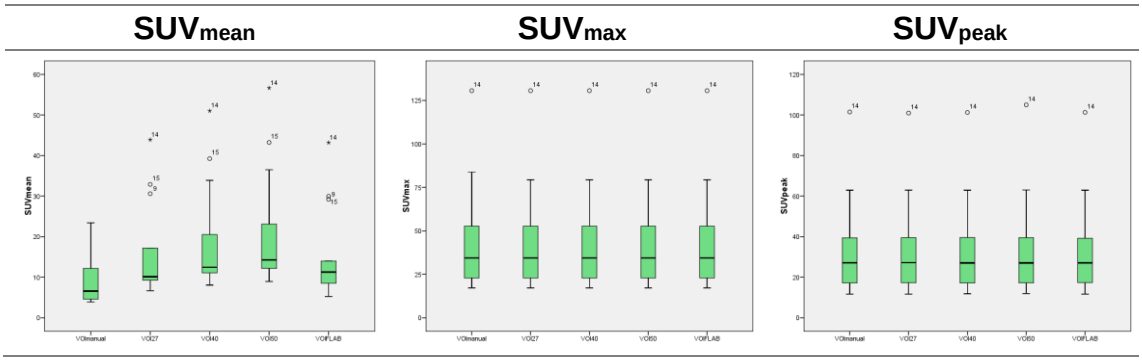


Figure 5-1: ^{18}F-Fluoride PET; Per-Patient Analysis; Baseline Scan; SUV Parameters; Descriptive statistics for all VOI methods and all patients - Boxplots

The choice of VOI definition technique affects the SUV parameters differently, with SUV_{mean} seeming to be more impacted than the other two parameters (see Tbl.5-1 and Fig.5-1).

SUV_{mean} increases as lower value voxels are excluded from the VOI by increasing the segmentation threshold. The lowest SUV_{mean} was calculated from VOI_{manual}; this reflects inclusion of marginal low value voxels. In addition, the range of SUV_{mean} is smallest using VOI_{manual}. The volume of the lower value voxels around the circumference of the VOI will have a greater impact on the overall SUV_{mean} than single higher intensity voxels in the middle of the VOI.

There is an absolute difference of 7.9 between the population average SUV_{mean} identified with VOI_{manual} compared to VOI_{50}; this represents a 234% difference. There is a maximum difference

of 131% (absolute SUV difference of 5.0) between the VOI techniques where an SUV threshold has been applied. This difference is statistically significant (a non-parametric Friedman's Variance test: where χ^2 = 53.2 (4 d.f.), p< 0.001). The distributions of SUV_{mean} calculated using VOI_{FLAB} is closest to the distributions achieved using VOI_{27}.

Clearly, when using $SUV_{mean,}$ the choice of thresholding technique will have a significant impact on the size of the calculated parameter.

SUV_{max} is more consistent, demonstrating smaller variations between the threshold methods. Although this was anticipated there are small differences. With application of a segmentation threshold, a single VOI can be separated into two or more smaller volumes the largest residual volume has been selected for analysis. This VOI may not contain the highest SUV_{max} voxel, thereby affecting the overall metric in this analysis. The differences identified in SUV_{max} across all the threshold techniques in these patients is not statistically significant (Friedman's Variance test, where χ^2 = 3 (4 d.f.), p=0.558).

A statistically significant difference is identified between the distributions of SUV_{peak} between the VOI (Friedman Test; where χ^2 =140.00 (4d.f.), *p*=0.007). The differences are small, with a maximum absolute difference of 0.66 (1.9%) between the average SUV_{peak} for each VOI threshold method. The methodology of this study has resulted in this impact on SUV_{peak} in the same way identified for SUV_{max}.

The distributions of the SUV parameters between the tumour groups have been compared using Mann-Whitney tests (see Tbl.5-2). There is no demonstrable statistically significant difference identified in these patients between the distributions of the SUV parameters tested of breast cancer and prostate cancer bone metastases, but there is a definite trend towards a higher SUV for the prostate cancer patients, particularly when the per-lesion analysis is considered which a significant difference. SUV_{peak} has more discriminating for these patients, but none of the parameters have met statistical significance. A larger sample would improve the statistical power of this analysis, and enable better comparison of the quantification parameters and VOI methods.

Prostate cancer metastases are usually more osteoblastic than those from breast cancer, and a higher ^{18}F-Fluoride uptake is physiologically expected.

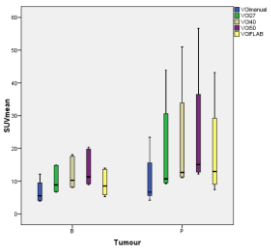
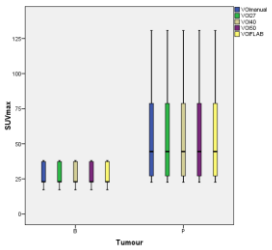
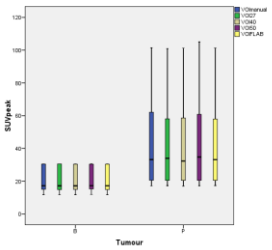
	SUV_{mean}		SUV_{max}		SUV_{peak}	
	Breast	**Prostate**	**Breast**	**Prostate**	**Breast**	**Prostate**
	Mean (Min-Max) [Median; SD]					
VOI_{Manual}	**7.05** (3.91-12.16) [5.55; 3.63]	**10.05** (4.2-23.4) [6.66; 6.69]	**27.65** (17.22-38.02) [22.92; 9.39]	**55** (22.69-130.63) [44.47; 36.41]	**20.98** (11.68-30.54) [17.22; 8.89]	**42.15** (17.14-101.48) [33.18; 28.3]
(p-value)	(0.298)		(0.112)		(0.083)	
VOI_{27}	**10.41** (6.68-14.88) [8.85; 4.15]	**19.21** (9.25-43.9) [10.69; 13.17]	**27.65** (17.22-38.02) [22.92; 9.39]	**54.51** (22.69-130.63) [44.47; 36]	**20.94** (11.66-30.54) [17.16; 8.96]	**41.76** (17.26-100.94) [33.95; 27.8]
p-value	(0.112)		(0.112)		(0.060)	
VOI_{40}	**12.44** (8.06-18.08) [10.28; 5]	**22.58** (11.06-51.01) [12.67; 15.04]	**27.65** (17.22-38.02) [22.92; 9.39]	**54.51** (22.69-130.63) [44.47; 36]	**20.95** (11.84-30.4) [17.17; 8.82]	**41.63** (17.04-101.24) [32.3; 28.01]
p-value	(0.112)		(0.112)		(0.083)	
VOI_{50}	**13.94** (8.96-20.45) [11.31; 5.71]	**25.05** (12.18-56.65) [15.08; 16.48]	**27.65** (17.22-38.02) [22.92; 9.39]	**54.51** (22.69-130.63) [44.47; 36]	**21.08** (11.95-30.56) [17.23; 8.78]	**42.58** (17.22-104.99) [34.71; 29.08]
p-value	(0.112)		(0.112)		(0.083)	
VOI_{FLAB}	**9.45** (5.27-13.99) [8.52; 4.16]	**18.16** (7.36-43.18) [12.95; 12.73]	**27.65** (17.22-38.02) [22.92; 9.39]	**54.58** (22.69-130.63) [44.47; 35.93]	**20.92** (11.66-30.51) [17.19; 8.91]	**41.66** (17.33-101.27) [33.15; 27.91]
p-value	(0.240)		(0.112)		(0.060)	
Box-plots						

Table 5-2: ^{18}F-Fluoride PET; Per-Patient Analysis; Baseline Scan; SUV Parameters; Comparison between tumour groups for all VOI methods and all patients (Mann-Whitney U tests for comparison)

5.1.2 Volumetric Parameters

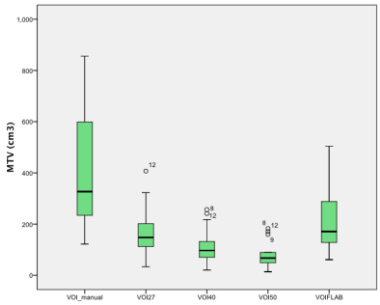
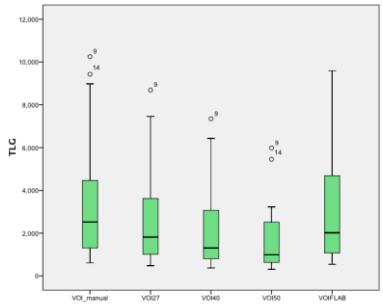
	MTV	TLA
	Mean (Min-Max) [Median; SD]	
VOI$_{Manual}$	**422.35** (122.05-855.84) [326.82; 249.49]	**3839.02** (606.14-10252.96) [2520.3; 3303.16]
VOI$_{27}$	**175.81** (33.84-407.05) [147.73; 102.22]	**2866.04** (478.95-8690.33) [1816.06; 2656.3]
VOI$_{40}$	**116.31** (20.44-256.64) [96.98; 72]	**2286.56** (371.62-7345.14) [1305.33; 2243.15]
VOI$_{50}$	**82.39** (14.08-182.59) [67.46; 51.9]	**1824.81** (305.3-5984.38) [989.24; 1846.87]
VOI$_{FLAB}$	**214.34** (61.22-504.04) [170.97; 135.58]	**3263.95** (543.49-9592.88) [2016.58; 2941.37]
Box Plots		

Table 5-3: ^{18}F-Fluoride PET; Per-Patient Analysis; Baseline Scan; Volumetric Parameters; Descriptive statistics for all VOI methods and all patients

The volumetric parameters will be affected by the threshold techniques; this is clearly demonstrated in the comparison of the MTV and TLA parameters (Tbl.5-3).

For MTV the differences in the distribution of the parameter across the threshold techniques is statistically significant (Friedman test (non-parametric): where χ^2=56.00, 4d.f., p<0.00001). The VOIs defined by the higher SUV thresholds are smaller because a larger volume of lower-value voxels are excluded from the final VOIs. However, the threshold used for VOI definition is of vital significance when calculating these volumetric parameters.

A similar effect is seen in the distributions of TLA, but to a smaller degree. TLA is the product of the MTV and SUV_{mean}. TLA is therefore also linked inextricably with the MTV, and the VOI methodology. This is confirmed using a Friedman Test (when χ^2=55.26, 4 d.f., p<0.00001). With increased severity of the threshold SUV_{max} used for VOI definition SUV_{mean} increases and MTV decreases; the effect on their product (TLA) is therefore blunted.

Table 5-4 gives an impression of the magnitude of the differences in the median of the distributions between the VOI definition methods. Large differences are shown in the distributions of the volumetric parameters depending on which VOI method is applied for calculating the parameters. They are essentially separate parameters – volumetric parameters without knowledge of the segmentation methods applied are not comparable. The distributions of the

volumetric parameters calculated with VOI_{FLAB} are most similar to the distributions calculated using VOI_{27}.

	VOI_{manual}/ VOI_{manual} %	VOI_{27}/ VOI_{manual} %	VOI_{40}/ VOI_{manual} %	VOI_{50}/ VOI_{manual} %	VOI_{FLAB}/ VOI_{manual} %
MTV	100	45	30	21	52
TLA	100	72	52	39	80

Table 5-4: ^{18}F-Fluoride PET; Per-Patient Analysis; Baseline Scan; Volumetric Parameters; Comparison (%) of median of distributions with VOI_{manual}

The volumetric parameters have been compared between the tumour groups in Table 5-5.

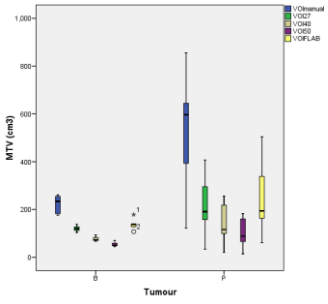
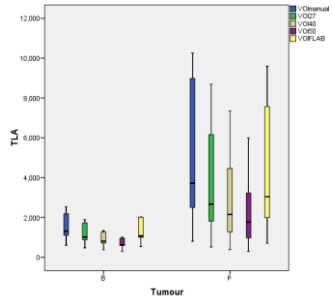
	MTV		**TLA**	
	Breast	**Prostate**	**Breast**	**Prostate**
	Mean (Min-Max) [Median; SD]			
VOI_{Manual}	**221.52** (176.03-260.53) [234.17; 39.96]	**533.91** (122.05-855.84) [596.66; 247.27]	**1548.32** (606.14-2537.69) [1305.36; 794.5]	**5111.63** (806.17-10252.96) [3720.87; 3508.98]
p-value	0.042		0.042	
VOI_{27}	**120.2** (103.76-137.89) [119.46; 13.24]	**206.7** (33.84-407.05) [190.95; 117.84]	**1198.1** (478.95-1892.98) [1014.77; 593.75]	**3792.68** (518.22-8690.33) [2666.46; 2929.98]
p-value	0.112		0.060	
VOI_{40}	**77.38** (66.7-94.37) [72.96; 11.16]	**137.94** (20.44-256.64) [116.08; 82.99]	**900.79** (371.62-1336.11) [807.28; 402.12]	**3056.42** (394.72-7345.14) [2152.06; 2495.74]
p-value	0.112		0.042	
VOI_{50}	**54.61** (45.43-69.83) [49.71; 9.87]	**97.82** (14.08-182.59) [88.04; 59.82]	**698.55** (307.06-1004.79) [631.26; 286.8]	**2450.51** (305.3-5984.38) [1763.94; 2066.01]
p-value	0.190		0.060	
VOI_{FLAB}	**138.18** (107.48-178.67) [137.82; 25.9]	**256.66** (61.22-504.04) [194.13; 154.58]	**1330.02** (543.49-2023.43) [1070.89; 658.76]	**4338.35** (704.51-9592.88) [3042.46; 3194.65]
p-value	0.147		0.060	
Box-plots				

Table 5-5: ^{18}F-Fluoride PET; Per-Patient Analysis; Baseline Scan; Volumetric Parameters; Comparison between tumour groups for all VOI methods and all patients (Mann-Whitney U tests for comparison)

The distributions of MTV and TLA between the tumour groups have been compared with Mann-Whitney U tests for each threshold technique pair. These results are shown in Table 5-5. For MTV, it is only the manual delineation with a significant difference between the tumour groups; this could suggest a difference in the penumbra of lower value pixels in the VOI, eliminated by applying a threshold relative to the SUV, and perhaps this is greater for prostate cancer

metastases. There is a trend for a larger MTV for the prostate patients using the other VOI methods, and this was identified at the per-lesion analysis, adding validity.

The TLA comparisons between breast and prostate cancer patients shows a likely significant difference irrespective of threshold technique. TLA is the product of MTV and SUV_{mean}; neither of these parameters have individually shown a difference between tumour types, and it is only the combination in a TLA returning significance. Both distributions of SUV_{mean} and MTV are empirically larger for these prostate cancer patients and this effect has been emphasised by the calculation of TLA. There is no VOI method showing superiority. The VOI_{27} iso-contour produces distributions of these SUV parameters most similar to the FLAB methodology.

At the per-patient analysis TLA seems more discriminatory when comparing the tumour groups, emphasising the higher baseline SUV_{mean} and MTV of the prostate lesions. The next chapter interrogates the potential predictive and prognostic role for these parameters.

5.1.3 Heterogeneity Parameters

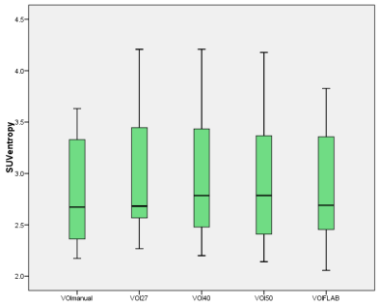
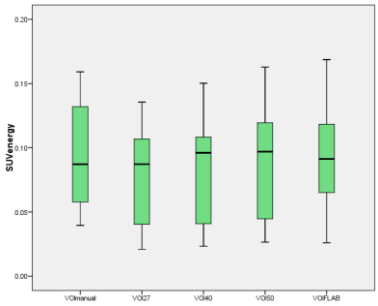
	$SUV_{entropy}$	SUV_{energy}
	Mean (Min-Max) [Median; SD]	
VOI_{Manual}	**2.8** (2.17-3.63) [2.67; 0.53]	**0.1** (0.04-0.16) [0.09; 0.04]
VOI_{27}	**2.97** (2.27-4.21) [2.68; 0.63]	**0.08** (0.02-0.14) [0.09; 0.04]
VOI_{40}	**2.94** (2.2-4.21) [2.78; 0.63]	**0.08** (0.02-0.15) [0.1; 0.04]
VOI_{50}	**2.92** (2.14-4.18) [2.79; 0.65]	**0.09** (0.03-0.16) [0.1; 0.04]
VOI_{FLAB}	**2.82** (2.06-3.83) [2.69; 0.58]	**0.1** (0.03-0.17) [0.09; 0.04]
Box Plots		

Table 5-6: ^{18}F-Fluoride PET; Per-Patient Analysis; Baseline Scan; SUV Heterogeneity Parameters; Descriptive statistics for all VOI methods and all patients

There is a small but significant difference in the distribution of $SUV_{entropy}$ between the threshold techniques (Friedman Test: where χ^2=17.13 (4 d.f.), p=0.002). This represents a maximum absolute difference of 0.17 between the mean $SUV_{entropy}$ measurements. The clinical significance of this size of difference is undetermined and will be interrogated in the next chapters. As the iso-contour threshold increases, entropy falls suggesting the VOI becomes more uniform as lower value voxels are excluded.

A significant difference also exists between the distributions of SUV_{energy} between the threshold methods. Using a parametric approach, where F=4.703, p=0.019 and η^2=0.299. There is a maximum difference of 0.02 between the means of these populations, but the clinical significance of a difference of this magnitude is undetermined.

The heterogeneity parameters seem much more consistent across the VOI methods than the other parameters. The per-lesion analysis demonstrated strong correlations between the SUV parameters and the SUV heterogeneity parameters; where the SUV parameters have shown inconsistencies across the VOI methods, the SUV heterogeneity parameters may record the same information in a method less reliant on the underlying VOI methodology.

The data suggests the distribution mean of $SUV_{entropy}$ is higher for prostate cancer bone metastases (see Tbl.5-7), but no statistically significant difference has been identified with non-parametric Mann Whitney U tests. No difference between the tumour groups has been shown

for SUV_{energy}. The per-lesion analysis also showed no difference in the SUV heterogeneity parameters between the tumour groups.

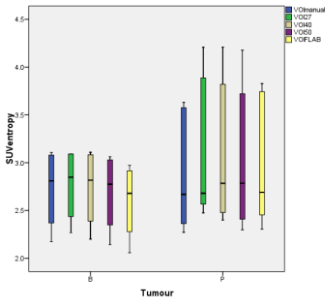
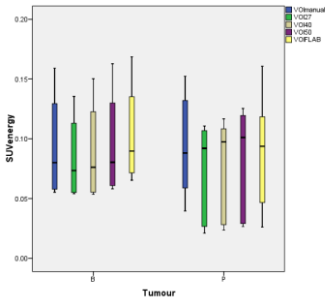
	$SUV_{entropy}$		**SUV_{energy}**	
	Breast	**Prostate**	**Breast**	**Prostate**
	Mean (Min-Max) [Median; SD]			
VOI_{Manual}	**2.62** (2.17-3.11) [2.57; 0.45]	**2.9** (2.27-3.63) [2.78; 0.57]	**0.1** (0.06-0.16) [0.1; 0.05]	**0.09** (0.04-0.15) [0.08; 0.04]
p-value	0.438		0.699	
VOI_{27}	**2.68** (2.27-3.09) [2.61; 0.4]	3.13 (2.48-4.21) [2.76; 0.69]	**0.09** (0.05-0.14) [0.09; 0.04]	**0.07** (0.02-0.11) [0.08; 0.04]
p-value	0.298		0.364	
VOI_{40}	**2.64** (2.2-3.11) [2.58; 0.43]	**3.11** (2.4-4.21) [2.95; 0.69]	**0.1** (0.05-0.15) [0.1; 0.04]	**0.07** (0.02-0.12) [0.1; 0.04]
p-value	0.240		0.438	
VOI_{50}	**2.69** (2.14-3.06) [2.78; 0.43]	**3.04** (2.3-4.18) [2.79; 0.74]	**0.1** (0.06-0.16) [0.08; 0.05]	**0.08** (0.03-0.13) [0.1; 0.04]
p-value	0.683		0.808	
VOI_{FLAB}	**2.51** (2.06-2.97) [2.5; 0.4]	**2.99** (2.31-3.83) [2.73; 0.62]	**0.11** (0.07-0.17) [0.1; 0.04]	**0.09** (0.03-0.16) [0.08; 0.04]
p-value	0.240		0.438	
Box Plots				

Table 5-7: ^{18}F-Fluoride PET; Per-Patient Analysis; Baseline Scan; SUV Heterogeneity Parameters; Comparison between tumour groups for all VOI methods and all patients (Mann-Whitney U tests for comparison)

5.1.4 Ki

n=13 – one patient had missing data and Ki could not be calculated.

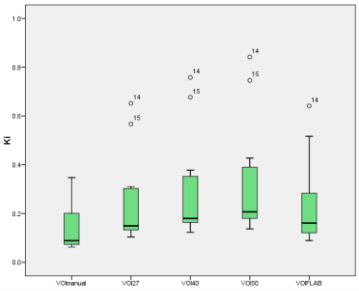
	Ki
	Mean (Min-Max) [Median; SD]
VOI_{Manual}	**0.13** (0.05-0.35) [0.09; 0.09]
VOI_{27}	**0.24** (0.09-0.65) [0.17; 0.15]
VOI_{40}	**0.28** (0.11-0.76) [0.2; 0.18]
VOI_{50}	**0.33** (0.14-0.84) [0.24; 0.21]
VOI_{FLAB}	**0.22** (0.07-0.64) [0.17; 0.15]
Box Plots	

Table 5-8: ^{18}F-Fluoride PET; Per-Patient Analysis; Baseline Scan; Ki; Descriptive statistics for all VOI methods and all patients

The calculation of Ki is affected by the choice of VOI method (Tbl.5-8). This can be demonstrated statistically (non-parametric Friedman Test: where χ^2 =45.267 (4 d.f.), p<0.00001). The tissue time-activity curve created for Patlak Ki analysis will be affected by the VOI threshold techniques – if only higher value voxels remain after segmentation, the average activity will be greater for a given VOI. The calculated Ki will therefore be higher. For this data, VOI_{FLAB} yields a Ki distribution most similar to VOI_{27}.

Comparison between the tumour groups has been performed (see Tbl.5-9). 13 patients had data for analysis, but only 12 with the VOI_{50} threshold. No statistically significant difference of the Ki between the tumour groups has been demonstrated (see table above). No difference was identified with the per-lesion analysis either (data not included for brevity).

It is not feasible from this analysis to identify a superior VOI methodology, but the differences in the quantified measurements are clear. The next chapter will review the clinical implications of the VOI methods.

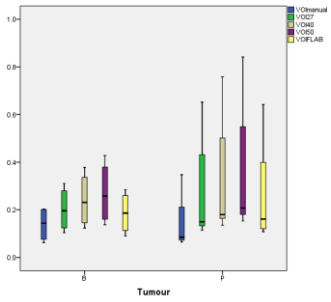
	Ki	
	Breast	**Prostate**
	Mean (Min-Max) [Median; SD]	
VOI_{Manual}	**0.12** (0.05-0.2) [0.09; 0.07]	**0.14** (0.06-0.35) [0.09; 0.1]
p-value	0.699	
VOI_{27}	**0.18** (0.09-0.31) [0.14; 0.1]	**0.27** (0.11-0.65) [0.15; 0.2]
p-value	0.438	
VOI_{40}	**0.21** (0.11-0.38) [0.17; 0.12]	**0.31** (0.14-0.76) [0.18; 0.24]
p-value	0.364	
VOI_{50}	**0.27** (0.14-0.43) [0.26; 0.13]	**0.36** (0.15-0.84) [0.21; 0.28]
p-value	0.683	
VOI_{FLAB}	**0.16** (0.07-0.29) [0.14; 0.09]	**0.25** (0.11-0.64) [0.15; 0.2]
p-value	0.438	
Box Plots		

Table 5-9: ^{18}F-Fluoride PET; Per-Patient Analysis; Baseline Scan; Ki; Comparison between tumour groups for all VOI methods and all patients (Mann-Whitney U tests for comparison)

5.2 Percentage Change (Δ%) Between Scans

12 patients had imaging pairs suitable for analysis of the percentage change (Δ%) between the baseline scan and the second scan.

5.2.1 Δ%SUV Parameters

The table below summaries the data for the SUV parameters (Tbl.5-10).

	Δ%SUV$_{mean}$	**Δ%SUV$_{max}$**	**Δ%SUV$_{peak}$**
		Mean Δ% (Min-Max) [Median; SD]	
VOI$_{Manual}$	**20.22** (-40.87-151.72) [7.88; 48.71]	**6.66** (-41.36-108.43) [1.07; 41.35]	**-0.13** (-43.62-94.64) [-1.99; 35.36]
VOI$_{27}$	**6.32** (-37.62-77.95) [-0.63; 34.44]	**1.67** (-48.43-96.21) [-3.98; 38.76]	**0.09** (-47.41-99.69) [-2.83; 37.42]
VOI$_{40}$	**6.47** (-37.44-70.67 [1.48; 32.76]	**0.51** (-53.94-125.85) [-12.43; 46.45]	**-1** (-49.81-104.69) [-4.78; 39.33]
VOI$_{50}$	**7.94** (-37.35-74.14 [3.43; 34.8]	**-1.12** (-56.92-140.3) [-17.99; 51.16]	**-2.24** (-52.49-104.41) [-7.18; 39.99]
VOI$_{FLAB}$	**9.74** (-35.61-84) [0.2; 39.2]	**2.61** (-44.08-87.44) [-1.48; 35.9]	**-0.96** (-47.21-91.98) [-2.44; 34.88]
VOI$_{27BL}$	**55.61** (-4.66-212.71) [20.62; 73.73]	**20.11** (-62.21-387.42) [-13.08; 120.38]	**-0.81** (-62.7-255.53) [-17.18; 84.27]
VOI$_{40BL}$	**57.26** (3.11-175.2) [30.73; 64.16]	**8.31** (-75.01-414.16) [-28.99; 131.62]	**-9.35** (-75.93-279.1) [-27.55; 94.25]
VOI$_{50BL}$	**59.66** (6.15-169.64) [38.62; 58.11]	**2.15** (-78.45-425.73) [-39.9; 136.56]	**-13.61** (-79.99-303.64) [-35.95; 102.72]

Table 5-10: ^{18}F-Fluoride PET; Per-Patient Analysis; %ΔSUV Parameters; Descriptive statistics for all VOI methods and all patients

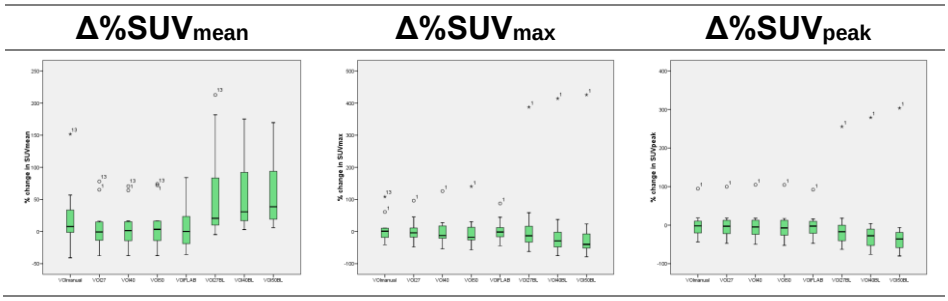


Figure 5-2: ^{18}F-Fluoride PET; Per-Patient Analysis; %ΔSUV Parameters; Descriptive statistics for all VOI methods and all patients – Box-Plots

I hypothesised the difference between the two scans might be consistent between the VOI methods applied; i.e. so long as there was consistency of the threshold method applied between scans, the change in parameter might be comparable between different techniques.

For SUV$_{mean}$ this hypothesis can be tested using a parametric one-way ANOVA. There is a statistically significant difference between the distributions of Δ%SUV$_{mean}$ between the VOI methods (when F=11.579, p=0.001, η^2=0.51). Much of this difference is caused by the Δ%SUV$_{mean}$ calculated using the absolute-SUV threshold methods for the second scan (VOI$_{27BL}$,

VOI_{40BL}, VOI_{50BL}). Absolute differences in the $\Delta\%SUV_{mean}$ from the VOI methods can be compared to those of $\Delta\%VOI_{manual}$ can through pairwise examination performed during the one-way ANOVA calculation (Tbl.5-11).

VOI Method	Absolute % difference in $\Delta\%SUV_{mean}$ compared to $\Delta\%VOI_{manual}$ (p-value)	Absolute % difference in $\Delta\%SUV_{max}$ compared to $\Delta\%VOI_{manual}$	Absolute % difference in $\Delta\%SUV_{peak}$ compared to $\Delta\%VOI_{manual}$
VOI_{27}	-13.9% (0.157)	-5.0%	+0.2%
VOI_{40}	-13.8% (0.169)	-6.2%	-0.9%
VOI_{50}	-12.3%(0.182)	-7.8%	-2.1%
VOI_{FLAB}	-10.5% (0.251)	-4.1%	-0.8%
VOI_{27BL}	35.4% (0.045)	13.5%	-0.7%
VOI_{40BL}	37.0% (0.033)	1.7%	-9.2%
VOI_{50BL}	39.4% (0.023)	-4.5%	-13.5%

Table 5-11: ^{18}F-Fluoride PET; Per-Patient Analysis; %ΔSUV Parameters; Difference between Δ% SUV Parameters by VOI method

There is a statistically significant difference between the distributions of $\Delta\%SUV_{mean}$ as calculated from the manual VOI technique and the SUV_{mean} calculated using VOIs defined by an absolute SUV threshold related to the baseline imaging. The differences recorded for the relative threshold methods are not statistically significant, but do demonstrate variation compared to manual delineation, with relative $\Delta\%SUV_{mean}$ values of between 31% and 50% of those of VOI_{manual}. In absolute terms, there is consistency between the $\Delta\%SUV_{mean}$ between the relative-SUV VOI methods, with only 3.4 percentage points difference between them.

The VOI methods based on the baseline scan SUV give dramatically different values for $\Delta\%SUV_{mean}$; for several cases a decline in SUV_{mean} by one threshold method corresponds to an increase in SUV_{mean} by another. This might affect the outcome of a concluded 'metabolic response' (MR); if a decrease in SUV_{mean} is used as a definition of a metabolic response, there were 11 'metabolic non-responders' (MNR) and only one MR with all relative VOI methods, but all patients were deemed to have an MR with the absolute SUV methods.

The differences between the distributions of $\Delta\%SUV_{max}$ from the threshold methods can be tested statistically, showing a real difference exists for these patients (Friedman Test: where χ^2=35.417 (7 d.f.), *p*=0.000009). However, the real differences are small between the relative-SUV methods used to define the VOI, with only 3.7 percentage points between the calculated values of $\Delta\%SUV_{max}$. A 13% difference is noted for $\Delta\%SUV_{max}^{50BL}$ compared to $\Delta\%SUV_{max}^{VOImanual}$.

A Friedman Test of the distributions of SUV_{peak} between the threshold techniques shows, where $\chi2$=44.472 (7 d.f.), p<0.00001. However, the relative SUV threshold methods are all within 2.3% of each other, and this is a small absolute difference in the calculated $\Delta\%SUV_{peak}$ and likely has little clinical significance. Again, the absolute SUV methods have produced values more different from $\Delta\%SUV_{peak}^{VOImanual}$ with a 13% difference noted compared with $\Delta\%SUV_{peak}^{VOI50BL}$.

5.2.2 Δ% Volumetric Parameters

The data for the %Δ volumetric parameters can be reviewed in Table 5-12.

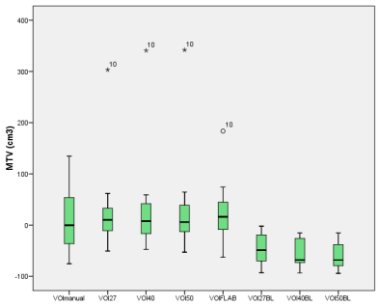
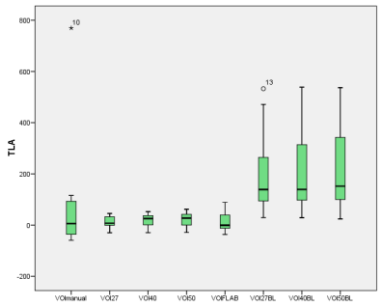
	Δ%MTV	Δ%TLA
	Mean (Min-Max) [Median; SD]	
VOI_{Manual}	**10.76** (-75.5-134.62) [-0.35; 64.82]	**75.93** (-59.21-769.62) [5.82; 226.41]
VOI_{27}	**30.09** (-50.64-303.25) [10.22; 91.39]	**12.6** (-29.96-45.54) [7.28; 22.94]
VOI_{40}	**33.4** (-47.4-340.91) [7.8; 101.96]	**19.22** (-29.41-52.42) [25.55; 24.77]
VOI_{50}	**32.46** (-52.89-342.02) [5.9; 103.23]	**21.73** (-28.53-61.91) [27.41; 26.7]
VOI_{FLAB}	**24.03** (-62.7-183.63) [16.29; 63.13]	**13.62** (-36.56-89.11) [-0.72; 40.17]
VOI_{27BL}	**-46.32** (-92.79--2.05) [-48.84; 28.93]	**199.97** (29.55-532.61) [139.18; 162.51]
VOI_{40BL}	**-55.86** (-93.55--15.31) [-68.02; 27.3]	**215.14** (29.03-538.71) [139.43; 173.66]
VOI_{50BL}	**-61.42** (-94.12--15.28) [-68.21; 26.84]	**219.46** (24.1-536.78) [151.98; 174.13]
Box Plots		

Table 5-12: ^{18}F-Fluoride PET; Per-Patient Analysis; %Δ Volumetric Parameters; Descriptive statistics for all VOI methods and all patients

The choice of threshold technique makes a significant difference in the measured change in MTV, identified using a one-way ANOVA (where F=10.693, p=0.002, η^2=0.493), but no significant difference between the relative threshold techniques is shown from pairwise comparisons (Tbl.5-13). The choice of threshold also affects the calculated TLA. Where F=4.334, p=0.057, η^2=0.283.

VOI Method	Absolute % difference in Δ%MTV compared to Δ%VOI_{manual} (p-value)	Absolute % difference in Δ%TLA compared to Δ%VOI_{manual}
VOI_{27}	19.33 (0.323)	9.13 (0.471)
VOI_{40}	22.64 (0.334)	9.86 (0.534)
VOI_{50}	21.71 (0.372)	8.85 (0.600)
VOI_{FLAB}	13.28 (0.239)	3.31 (0.352)
VOI_{27BL}	-57.08 (0.008)	-62.61 (0.064)
VOI_{40BL}	-66.62 (0.003)	-71.36 (0.035)
VOI_{50BL}	-72.17 (0.001)	-77.03 (0.023)

Table 5-13: ^{18}F-Fluoride PET; Per-Patient Analysis; %Δ Volumetric Parameters; Pairwise comparison of distributions between VOI methods

5.2.3 Δ% SUV Heterogeneity Parameters

The data for the %Δ SUV Heterogeneity parameters can be reviewed in Table 5-14.

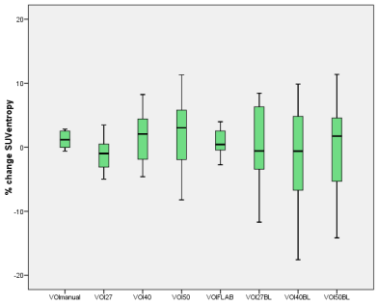
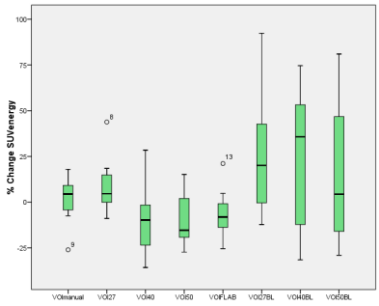
	Δ%$SUV_{entropy}$	Δ%SUV_{energy}
	Mean (Min-Max) [Median; SD]	
VOI_{Manual} n=12	**-0.71** (-11.85-2.83) [0.46; 4.08]	**1.53** (-26.03-59.48) [1.74; 23.78]
VOI_{2} n=12	**-1.51** (-11.42-3.48) [-0.73; 4.06]	**0.54** (-59.83-45.57) [-0.05; 29.44]
VOI_{40} n=10	**0.09** (-14.93-8.24) [1.79; 6.48]	**-11 00** (-35.71-28.41) [-11.2; 17.8]
VOI_{50} n=10	**2.01** (-12.18-16.13) [2.63; 8.55]	**-15.89** (-48.76-15.16) [-18.12; 19.02]
VOI_{FLAB} n=12	**0.64** (-2.73-3.99) [0.34; 1.87]	**-14.04** (-63.73-21.1) [-9.05; 21.31]
VOI_{27BL} n=10	**0.57** (-12.59-9.19) [2.78; 8.2]	**14.21** (-30.5-92.36) [3.03; 37.16]
VOI_{40BL} n=9	**-0.8** (-17.56-15.88) [-0.6; 10.67]	**24.55** (-31.52-85.46) [35.75; 45.93]
VOI_{50BL} n=7	**-0.36** (-14.15-11.38) [1.74; 8.56]	**16.86** (-29.01-81) [4.37; 41.59]
Box Plots		

Table 5-14: ^{18}F-Fluoride PET; Per-Patient Analysis; %ΔSUV Heterogeneity Parameters; Descriptive statistics for all VOI methods and all patients

The distribution of $\%\Delta SUV_{entropy}$ is not significantly different across the threshold methods (one-way ANOVA, where F=0.562, p=0.622, η^2=0.086). The differences, although not statistically significant, are small. When comparing the absolute differences in the mean of the distributions with $VOI_{manual,}$ all threshold techniques produce a result within 2.72% of each other (Tbl.5-15).

The distributions of $\Delta\%$ SUV_{energy} between the threshold methods approaches statistical significance with a non-parametric analysis (Friedman Test: when χ^2 =13.095 (7 d.f.), p=0.07.

	Absolute % difference in mean $\Delta\%SUV_{entropy}$ compared to mean $\Delta\%VOI_{manual}$ (p-value)	**Absolute % difference in mean $\Delta\%SUV_{energy}$ compared to mean $\Delta\%VOI_{manual}$ (p-value)**
VOI_{27}	-2.30 (0.072)	8.97 (0.29)
VOI_{40}	0.34 (0.844)	-10.5 (0.28)
VOI_{50}	0.78 (0.743)	-9.73 (0.18)
VOI_{FLAB}	-0.36 (0.753)	-6.84 (0.39)
VOI_{27BL}	-0.92 (0.760)	25.52 (0.09)
VOI_{40BL}	-2.93 (0.464)	22.09 (0.26)
VOI_{50BL}	-1.57 (0.646)	15.96 (0.44)

Table 5-15: ^{18}F-Fluoride PET; Per-Patient Analysis; %Δ SUV Heterogeneity Parameters; Pairwise comparison of distributions between VOI methods

5.2.4 Δ% Ki

There is a wide spread of measured %ΔKi depending on the threshold methods used (see Tbl.5-16). These differences are not statistically significant (Friedman Test: where χ^2 =12.714 (7 d.f.), p=0.079); for this small number of patients the differences between the distributions are not large enough to be quite statistically significant (see Tbl.5-17). The largest differences are from the absolute SUV threshold methods, but statistical significance is not reached, and therefore it is not possible to draw firm conclusions from this.

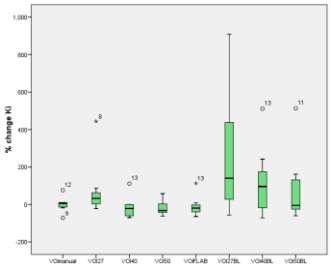
	Δ%Ki
	Mean (Min-Max) [Median; SD]
VOI_{Manual} **n=12**	**34.75** (-71.45-508.97) [-1.03; 154]
VOI_{2} **n=12**	**53.23** (-88.55-443.97) [5.08; 144.56]
VOI_{40} **n=10**	**-16.77** (-69.72-111.12) [-26.47; 55.76]
VOI_{50} **n=10**	**-30.54** (-83.8-57.86) [-39.43; 43.19]
VOI_{FLAB} **n=12**	**-23.38** (-90.83-113.14) [-25.52; 50.89]
VOI_{27BL} **n=10**	**185.61** (-56.7-909.73) [36.97; 327.11]
VOI_{40BL} **n=9**	**115.06** (-71.68-511.21) [95.97; 191.13]
VOI_{50BL} **n=7**	**94.47** (-60.18-513.25) [-4.55; 201.21]
Box Plot	

Table 5-16: ^{18}F-Fluoride PET; Per-Patient Analysis; %ΔKi; Descriptive statistics for all VOI methods and all patients

	Absolute % difference in Δ%Ki compared to $Δ\%VOI_{manual}$ (p-value)
VOI_{27}	-2.3 (0.07)
VOI_{40}	0.34 (0.84)
VOI_{50}	0.78 (0.74)
VOI_{FLAB}	-0.36 (0.75)
VOI_{27BL}	-0.92 (0.76)
VOI_{40BL}	-2.93 (0.46)
VOI_{50BL}	-1.57 (0.65)

Table 5-17: ^{18}F-Fluoride PET; Per-Patient Analysis; %ΔKi; Pairwise comparison of distributions between VOI methods

5.3 Discussion

	SUV_{mean}	SUV_{max}	SUV_{peak}	MTV	TLA	$SUV_{entropy}$	SUV_{energy}	Ki
Tumour	Prostate >Breast	Prostate >Breast	Prostate >Breast	Prostate >Breast	Prostate >Breast	X	X	X
VOI Superiority	X	X	X	X	X	X	X	X

Table 5-18: ^{18}F-Fluoride PET Per-Patient analysis; Descriptive statistics; Baseline Scan; Summary Table

	$\%\Delta SUV_{mean}$	$\%\Delta SUV_{max}$	$\%\Delta SUV_{peak}$	%ΔMTV	%ΔTLA	$\%\Delta SUV_{entropy}$	$\%\Delta SUV_{energy}$	%ΔKi
Tumour	X	X	X	X	X	X	X	X
VOI Superiority	X	X	X	X	X	X	X	X

Table 5-19: ^{18}F-Fluoride PET Per-Patient analysis; Descriptive statistics; %Δ parameters; Summary Table

There is little consistency of PET quantification methodology in the literature. Comparison of studies is limited due to these inconsistencies, but there is sparse evidence to document how the methodologies of quantification impact the resulting measurements and their clinical utility.

Erdi et al. used phantom studies to interrogate the impact of VOI iso-contour methodology using ^{18}F-FDG PET imaging, showing a fixed threshold between 35-44% of SUV_{max} had most correlation with the known volume of disease, but only for target spheres greater than 4mls[107]. Low concordance of tumour volumes has been demonstrated when different VOI threshold definition methods are used, with up to 41% difference in measured volumes identified [101,100] . Tomasi et al. reported the reproducibility of tumour volume using ^{11}C-Choline PET iso-contour techniques for definition of prostate gland tumours, including fixed and adaptive iso-contour methods, and FLAB. The reproducibility of the parameters was best using FLAB with the iso-contour methods resulting in the least satisfactory results[207].

Phantom studies do not necessarily translate into the much more complex tissue associations in-vivo. There are few studies correlating the PET volume with the histological volume (e.g.[208,209]). To my knowledge there is no literature providing histological confirmation of PET delineation of bone metastases, or examining the impact of threshold methodology on the semi-quantification parameters.

This study has demonstrated the varying impact of iso-contour and FLAB segmentation methods on a selection of semi-quantification parameters of bone metastases using ^{18}F-Fluoride PET. As a higher threshold for segmentation is applied, more lower value voxels are eliminated from the periphery of the VOI, resulting in an increase of SUV_{mean}. The segmentation algorithm was applied only to the edge of the and the centre of the VOI may still contain lower value voxels. The reduction in volume, and increase in SUV_{mean} with increasing threshold severity suggests the VOIs have a more avid centre, with a decreasing SUV gradient to the tumour edge; this is not

uniform as some VOIs were split into two more distinct VOIs by increasing iso-contour thresholds. The SUV_{mean} varied by up to 8 between the methods. The SUV_{max} and SUV_{peak} measurements seem more robust. Small differences are identified because of the higher thresholds separating VOIs into separate volumes - the largest (not the 'hottest') was used for quantification.

The SUV heterogeneity parameters are closely and strongly correlated with the SUV parameters, but they are much more consistent across the VOI methods than SUV_{mean}. SUV_{mean} and the heterogeneity parameters describe the whole VOI voxel population, rather than just the selection measured with SUV_{max} and SUV_{peak}, and therefore may have clinical relevance; despite the strong and significant correlation with the SUV parameters, the lower impact of the VOI method on the heterogeneity parameters may suggest more robust parameters.

For these lesions, the 27% iso-contour correlated the PET volume closest to the sclerotic component of the bone metastasis on the CT component of the scan. The FLAB methodology has produced baseline parameters most similar to the VOI_{27} methodology, suggesting FLAB may offer a robust semi-automated segmentation technique not reliant on the SUV_{max} and with favourable repeatability[99].

The %Δ parameters shows less variation between the VOI methods than analysis of the raw data alone. The SUV_{mean} data shows only 3.4% difference in the %Δ SUV_{mean} measurement between the relative-SUV VOI methods. This provides reassurance when comparing %Δ parameters between studies, provided there has been consistency of the VOI methodology between scans. The absolute-SUV VOI methods, however produce significantly different distributions of the %Δ parameters, and are not directly comparable with the other VOI methods. The choice of VOI method is clearly of vital significance; the direction of change of a parameter, in addition to the magnitude of change, is directly impacted by the VOI method, and, depending on the definition of metabolic response, these differences might be the difference between identifying a patient as a treatment responder or not. The $\%\Delta SUV_{entropy}$ parameter seems much less impacted by the VOI method than other parameters.

It has not been feasible to identify the best VOI method from this analysis. What is clear, however, is the choice of VOI definition method has a significant and large impact on the quantification parameters. This impact is different depending on the parameter. The most widely used parameters of SUV_{max} and SUV_{peak} are less influenced by the VOI method, and this gives them a practical advantage. The other parameters, however, including SUV_{mean}, volumetric parameters, heterogeneity parameters and Ki are all significantly impacted by the choice of segmentation methodology. An important area for further investigation would be the repeatability and reproducibility of the parameters as the VOI methodologies may perform differently when considering these. The ground truth of where the tumour tissue starts and stops is not as clinically relevant with the PET parameters as the clinical information gained from quantification. If a quantification method can be shown to be reproducible and repeatable, and has good diagnostic, predictive or prognostic accuracy, then the histological truth is perhaps less relevant.

Chapter 6 ^{18}F-Fluoride Per-Patient Analysis - Response analysis, and the impact of VOI segmentation methods

The methods of this chapter and the statistical approaches for this section can be reviewed in Chapter 2.3.5.

6.1 Baseline Scan

12 patients (5 with breast cancer (3 with PD) , 7 with prostate cancer (6 with PD)) had 24-week response data for analysis. Treatment response rate, therefore, for breast cancer = 60%, for prostate cancer=14%.

6.1.1 SUV Parameters

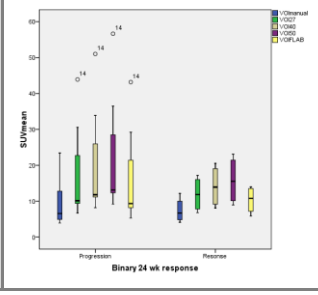
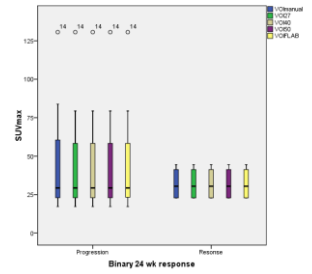
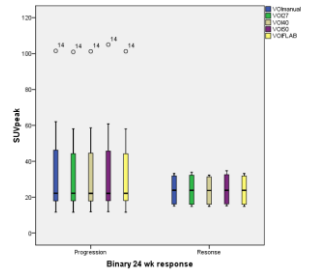
	SUV_{mean}		**SUV_{max}**		**SUV_{peak}**	
	Progression n=8	**Response n=4**	**Progression n=8**	**Response n=4**	**Progression n=8**	**Response n=4**
	Mean (Min-Max) [Median; SD]					
VOI_{Manual}	**9.48** (3.91-23.4) [6.57; 6.84]	**7.41** (4.12-12.16) [6.69; 3.51]	**46.69** (17.22-130.63) [29.34; 39.85]	**32.07** (22.87-44.47) [30.47; 10.92]	**35.74** (11.68-101.48) [22.2; 30.77]	**24** (15.08-33.18) [23.88; 9.18]
p-value	0.808		1.00		0.933	
VOI_{27}	**16.88** (6.68-43.9) [10.13; 13.26]	**11.93** (6.81-17.17) [11.87; 4.89]	**46.14** (17.22-130.63) [29.34; 39.29]	**32.07** (22.87-44.47) [30.47; 10.92]	**35.2** (11.66-100.94) [22.2; 30.14]	**24.12** (14.84-33.95) [23.85; 9.53]
p-value	0.808		1.00		0.808	
VOI_{40}	**19.63** (8.19-51.01) [11.8; 15.06]	**14.11** (8.06-20.53) [13.93; 5.9]	**46.14** (17.22-130.63) [29.34; 39.29]	**32.07** (22.87-44.47) [30.47; 10.92]	**35.27** (11.84-101.24) [22.13; 30.29]	**23.7** (14.95-32.3) [23.77; 8.9]
p-value	0.570		1.00		0.808	
VOI_{50}	**21.71** (9.2-56.65) [13.11; 16.57]	**15.78** (8.96-23.09) [15.54; 6.73]	**46.14** (17.22-130.63) [29.34; 39.29]	**32.07** (22.87-44.47) [30.47; 10.92]	**36.07** (11.95-104.99) [22.11; 31.67]	**24.4** (15.26-34.71) [23.81; 9.61]
p-value	0.570		1.00		0.808	
VOI_{FLAB}	**15.8** (5.27-43.18) [9.34; 13.34]	**10.35** (5.86-13.99) [10.79; 3.83]	**46.22** (17.22-130.63) [29.34; 39.24]	**32.07** (22.87-44.47) [30.47; 10.92]	**35.21** (11.66-101.27) [22.14; 30.23]	**23.93** (14.88-33.15) [23.85; 9.23]
p-value	0.808		1.00		0.808	
Box-plots						

Table 6-1: ^{18}F-Fluoride PET Per-Patient Analysis; Baseline SUV Parameters; Response Analysis (Mann-Whitney U tests for comparison)

The distribution of the SUV parameters measured seem higher for the patients with PD, but no significant difference is identified with Mann-Whitney U tests (Tbl.6-1). There is no significant difference when the tumour groups are considered independently, although with such small numbers statistical analysis is limited (results not included for brevity).

No association between the baseline SUV parameters and treatment response was confirmed with per-lesion analysis either. The highest SUVs were in patients with PD, and this study may be underpowered to identify the difference.

6.1.2 Volumetric Parameters

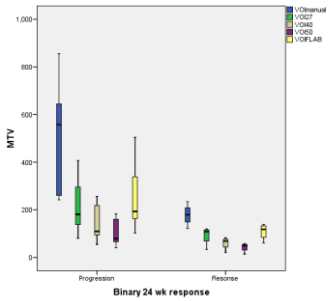
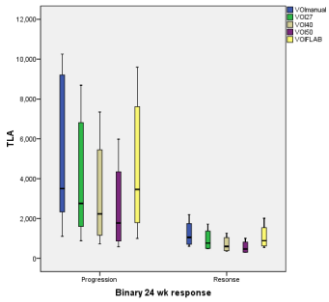
	MTV (cm^3)		TLA	
	Progression n=8	**Response n=4**	**Progression n=8**	**Response n=4**
	Mean (Min-Max) [Median; SD]			
VOI_{Manual}	**515.34** (254.61-855.84) [557.66; 204.96]	**178.63** (122.05-234.17) [179.15; 45.85]	**5179.97** (1107.6-10252.96) [3501.4; 3753.33]	**1225.62** (606.14-2184.81) [1055.77; 703.82]
p-value	0.004		0.028	
VOI27	**227.51** (127.6-407.05) [186.03; 101.9]	**92.34** (33.84-119.46) [108.03; 39.52]	**3988.97** (880.59-8690.33) [2757.9; 3036.51]	**933.78** (478.95-1723.19) [766.49; 580.02]
p-value	0.004		0.028	
VOI_{40}	**153.83** (72.96-256.64) [123.91; 72.96]	**60** (20.44-82.52) [68.52; 27.23]	**3217.79** (719.73-7345.14) [2229.59; 2590.19]	**710.71** (371.62-1269.23) [601; 422.73]
p-value	0.008		0.028	
VOI_{50}	**109.35** (45.43-182.59) [88.66; 53.2]	**42.97** (14.08-59) [49.4; 19.78]	**2570.16** (593.43-5984.38) [1777.4; 2153.41]	**562.1** (305.3-1004.79) [469.16; 332.54]
p-value	0.028		0.048	
VOI_{FLAB}	**283.68** (138.75-504.04) [240.4; 141.14]	**108.68** (61.22-137.82) [117.83; 34.08]	**4541.4** (1002.57-9592.88) [3461.2; 3325.62]	**1082.16** (543.49-2009.74) [887.7; 656.59]
p-value	0.004		0.028	
Box-plots				

Table 6-2: ^{18}F-Fluoride PET Per-Patient Analysis; Baseline Volumetric Parameters; Response Analysis; (Mann-Whitney U tests for comparison)

The data (Tbl.6-2) confirms significantly higher MTV and TLA distributions in those with subsequent PD. This was also identified at the per-lesion analysis, but MTV and TLA were not independent of the tumour groups. There are too few patients to perform sub-group analysis. Independence of these volumetric parameters as response predictors cannot be concluded.

6.1.3 Heterogeneity Parameters

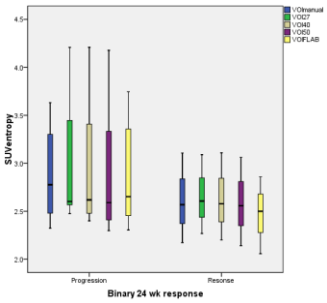
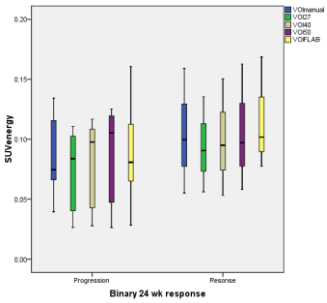
	$SUV_{entropy}$		SUV_{energy}	
	Progression n=8	**Response n=4**	**Progression n=8**	**Response n=4**
	Mean (Min-Max) [Median; SD]			
VOI_{Manual}	**2.81** (2.21-3.63) [2.67; 0.55]	**2.7** (2.17-3.11) [2.76; 0.42]	**0.0943** (0.0396-0.1423) [0.0881; 0.0383]	**0.0983** (0.0553-0.1591) [0.0894; 0.0444]
p-value	0.933		1.00	
VOI_{27}	**2.96** (2.34-4.21) [2.6; 0.69]	**2.79** (2.27-3.2) [2.85; 0.43]	**0.0777** (0.0266-0.1145) [0.092; 0.0368]	**0.0844** (0.0552-0.1354) [0.0734; 0.0378]
p-value	1.00		0.808	
VOI_{40}	**2.91** (2.27-4.21) [2.56; 0.71]	**2.77** (2.2-3.18) [2.84; 0.46]	**0.0831** (0.0279-0.1199) [0.1023; 0.0391]	**0.0886** (0.0535-0.1503) [0.0753; 0.0454]
p-value	1.00		0.933	
VOI_{50}	**2.94** (2.3-4.18) [2.59; 0.73]	**2.59** (2.14-3.06) [2.56; 0.46]	**0.0844** (0.0265-0.1253) [0.1054; 0.0432]	**0.106** (0.0581-0.1628) [0.0971; 0.0529]
p-value	0.667		0.833	
VOI_{FLAB}	**2.81** (2.18-3.74) [2.57; 0.62]	**2.6** (2.06-2.99) [2.68; 0.42]	**0.0966** (0.0283-0.1607) [0.0937; 0.0451]	**0.1062** (0.0769-0.1686) [0.0897; 0.0432]
p-value	0.933		0.808	
Box-plots				

Table 6-3: ^{18}F-Fluoride PET Per-Patient Analysis; Baseline SUV Heterogeneity Parameters; Response Analysis; (Mann-Whitney U tests for comparison)

The $SUV_{entropy}$ distributions (Tbl.6-3) appear to be higher in those with PD, and the SUV_{energy} distributions lower, but these differences are not statistically significant (Mann-Whitney U tests, Tbl.6-3). The sample size will limit identification of a smaller difference, and sub-group analysis is not feasible. Per-lesion analysis of these parameters also did not suggest response prediction utility.

6.1.4 Ki

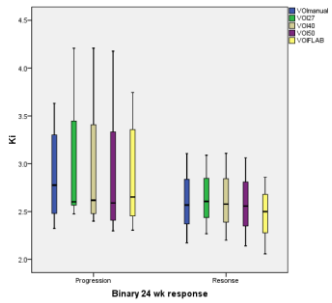
	Ki	
	Progression n=8	**Response n=4**
	Mean (Min-Max) [Median; SD]	
VOI_{Manual}	**2.81** (2.21-3.63) [2.67; 0.55]	**0.11** (0.06-0.2) [0.09; 0.06]
P-value	0.933	
VOI_{27}	**2.96** (2.34-4.21) [2.6; 0.69]	**0.18** (0.1-0.25) [0.17; 0.06]
P-value	1.00	
VOI_{40}	**2.91** (2.27-4.21) [2.56; 0.71]	**0.21** (0.12-0.29) [0.21; 0.08]
P-value	1.00	
VOI_{50}	**2.94** (2.3-4.18) [2.59; 0.73]	**0.22** (0.14-0.33) [0.18; 0.1]
P-value	0.667	
VOI_{FLAB}	**2.81** (2.18-3.74) [2.57; 0.62]	**0.15** (0.09-0.23) [0.15; 0.06]
P-value	0.933	
Box-plots		

Table 6-4: ^{18}F-Fluoride PET Per-Patient Analysis; Baseline Ki; Response Analysis; (Mann-Whitney U tests for comparison)

No statistically significant difference has been identified in the distribution of Ki between the two response groups (Tbl.6-4) despite the box-plot suggesting a higher Ki in the those with PD, but the sample number is small. There was no demonstrable difference between the tumour subgroups, but the highest values were from prostate cancer patients. Tumour subgroup analysis is not feasible for this data, but it is likely Ki is not an independent predictor of response. The per-lesion analysis also did not show predictive utility of the Ki parameter for these patients.

6.2 Percentage Change (%Δ) Between Scans

n=11 (7 with PD, 4 with a treatment response)

6.2.1 %Δ SUV Parameters

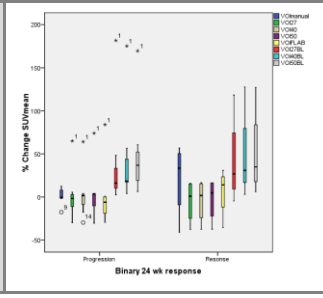
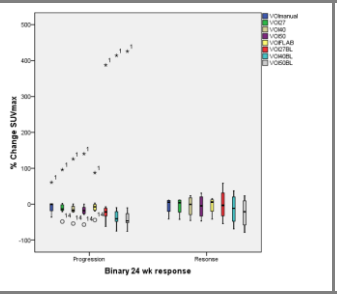
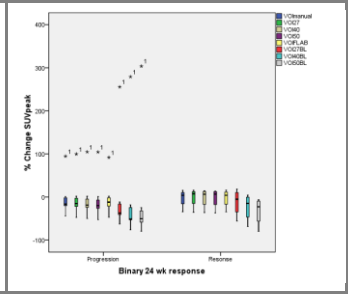
	%ΔSUV$_{mean}$		%ΔSUV$_{max}$		%ΔSUV$_{peak}$	
	Progression n=7	**Response n=4**	**Progression n=7**	**Response n=4**	**Progression n=7**	**Response n=4**
	Mean (Min-Max) [Median; SD]					
VOI$_{Manual}$	**1.13** (-17.55-12.53) [-0.33; 10.04]	**20.75** (-40.87-56.86) [33.5; 43.39]	**-1.32** (-35.67-60.95) [-0.37; 30.66]	**-4.81** (-41.36-10.99) [5.56; 24.76]	**-1.39** (-43.62-94.64) [-16.61; 44.65]	**-2.65** (-34.02-15.33) [4.05; 21.61]
p-value	0.230		0.527		0.412	
VOI$_{27}$	**2.48** (-29.87-65.07) [-1.67; 30.01]	**-4.86** (-37.62-16.14) [1.01; 25.23]	**-0.51** (-48.43-96.21) [-13.51; 45.56]	**-5.46** (-42.18-11.98) [4.19; 25.28]	**-1.76** (-47.41-99.69) [-15.07; 47.29]	**-1.27** (-35.53-15.46) [7.5; 23.19]
p-value	1.00		0.527		0.412	
VOI$_{40}$	**3.48** (-29.64-64.12) [1.42; 29.56]	**-4.34** (-37.44-16.48) [1.79; 25.26]	**0.19** (-53.94-125.85) [-16.87; 57.82]	**-5.71** (-45.35-23.76) [-0.64; 30.43]	**-3.03** (-49.81-104.69) [-18.72; 50]	**-2.18** (-36.11-14.45) [6.48; 23.23]
p-value	1.00		0.648		0.412	
VOI$_{50}$	**4.96** (-30.26-74.14) [3.24; 33.49]	**-2.88** (-37.35-16.42) [4.7; 25.45]	**-0.67** (-56.92-140.3) [-18.07; 64.47]	**-6.28** (-46.68-30.66) [-4.56; 33.85]	**-4.97** (-52.49-104.41) [-20.06; 50.82]	**-2.39** (-36.81-13.74) [6.76; 23.5]
p-value	1.00		0.788		0.527	
VOI$_{FLAB}$	**1.61** (-29.39-84) [-6.01; 37.99]	**5.9** (-35.61-30.98) [14.12; 28.84]	**0.22** (-44.08-87.44) [-7.37; 41.29]	**-3.66** (-41.57-15.28) [5.83; 25.84]	**-1.91** (-47.21-91.98) [-12.01; 44.22]	**-2.64** (-34.38-15.67) [4.07; 22.01]
p-value	0.527		0.527		0.527	
VOI$_{27BL}$	**41.06** (2.5-181.74) [16.12; 63.74]	**41.81** (-4.66-118.34) [26.78; 53.22]	**30.69** (-62.21-387.42) [-21.59; 158.28]	**-0.55** (-54.42-58.54) [-3.15; 46.6]	**5.96** (-62.7-255.53) [-37.43; 111.24]	**-12.04** (-55.87-17.92) [-5.1; 31.85]
p-value	0.788		0.527		0.527	
VOI$_{40BL}$	**45.66** (4.06-175.2) [18.52; 59.47]	**48.16** (3.11-127.67) [30.93; 54.61]	**23.05** (-75.01-414.16) [-40.66; 173.58]	**-13.63** (-68.81-37.47) [-11.59; 45.1]	**-0.43** (-75.93-279.1) [-51.37; 124.61]	**-23.5** (-68.67-4.35) [-14.83; 32.36]
p-value	1.00		0.527		0.527	
VOI$_{50BL}$	**50.85** (6.53-169.64) [36.83; 55.48]	**50.77** (6.15-127.05) [34.94; 52.82]	**21.08** (-76.13-425.73) [-46.42; 179.5]	**-24.21** (-78.45-23.66) [-21.02; 43.92]	**-1.02** (-79.25-303.64) [-50.4; 135.42]	**-32.89** (-79.99--6.31) [-22.63; 33.23]
p-value	0.927		0.648		0.648	
Box-plots						

Table 6-5: ^{18}F-Fluoride PET Per-Patient Analysis; %Δ SUV Parameters; Response Analysis; (Mann-Whitney U tests for comparison)

No statistically significant difference in %ΔSUV$_{mean}$, %ΔSUV$_{max}$, or %ΔSUV$_{peak}$ has been identified between the 24-week response groups, irrespective of the threshold technique used when compared with Mann-Whitney tests (Tbl.6-5).

The per-lesion analysis showed an increase in SUV between the scans was associated with a treatment response. It is not feasible to conduct tumour-subgroup analysis at the per-patient

level. This analysis is underpowered due to the small number of patients, and therefore the lack of clinical utility of these parameters cannot be excluded.

No VOI method has obvious superiority from this analysis.

6.2.2 %Δ Volumetric Parameters

Table 6-6 reports comparison of the %Δ volumetric parameters between the 24-week response groups, comparing the groups with Mann-Whitney U tests for each VOI method. None of the analyses approach statistical significance. Tumour subgroup analyses are not feasible given the small number of patients in each tumour group. The per-lesion analysis also did not show predictive utility of the %Δ volumetric parameters. No VOI method has shown superiority.

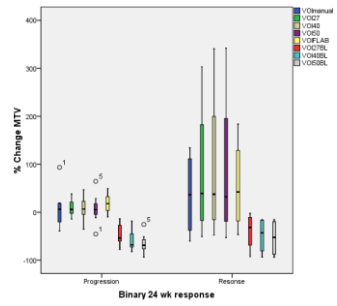
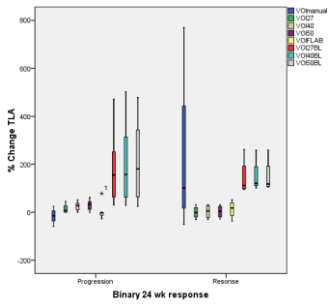
	%ΔMTV		%ΔTLA	
	Progression n=7	**Response n=4**	**Progression n=7**	**Response n=4**
	Mean (Min-Max) [Median; SD]			
VOI_{Manual}	**8.14** (-39.3-93.51) [6.19; 44.17]	**36.89** (-60.14-134.62) [36.54; 89.98]	**-15.74** (-59.21-24.27) [-14.65; 30.78]	**230.29** (-50.06-769.62) [100.8; 366.72]
p-value	1.00		0.164	
VOI_{27}	**9.65** (-14.11-38.03) [5.83; 18.63]	**82.7** (-50.64-303.25) [39.09; 154.12]	**15.29** (-0.72-45.54) [7.86; 18.19]	**-0.16** (-29.96-32.1) [-1.4; 26.22]
p-value	0.412		0.315	
VOI_{40}	**7.85** (-35.34-46.92) [6.56; 27.66]	**92.15** (-47.4-340.91) [37.54; 171.51]	**24.77** (-0.08-52.42) [27.65; 19.46]	**2.99** (-29.41-31.2) [5.09; 29.03]
p-value	0.412		0.230	
VOI_{50}	**7.19** (-45.52-64.48) [5.31; 33.9]	**88.3** (-52.89-342.02) [32.04; 174.37]	**29.59** (-0.66-61.91) [32.89; 23.13]	**2.31** (-28.53-30.4) [3.69; 26.93]
p-value	0.527		0.109	
VOI_{FLAB}	**18.5** (-9.74-49.3) [17.91; 22.12]	**55.4** (-46.59-183.63) [42.29; 98.7]	**3.29** (-27.16-78.37) [-2.44; 34.43]	**12.83** (-36.56-52.15) [17.87; 37.37]
p-value	0.788		0.527	
VOI_{27BL}	**-45.69** (-77.73--13.41) [-53.78; 23.76]	**-39.6** (-92.79--2.05) [-31.77; 39.39]	**183.94** (29.55-471.15) [155.43; 156.68]	**144.87** (94.71-261.55) [111.61; 78.74]
p-value	0.788		1.00	
VOI_{40BL}	**-57.64** (-82.2--18.46) [-68.09; 23.34]	**-48.71** (-93.55--15.31) [-42.98; 38.46]	**205.89** (29.03-502.63) [157.35; 179.76]	**150.45** (101.69-258.5) [120.8; 72.6]
p-value	0.527		1.00	
VOI_{50BL}	**-65.02** (-93.72--25.46) [-68.91; 21.84]	**-53.58** (-94.12--15.28) [-52.47; 39.53]	**213.96** (24.1-479.25) [180.38; 181.29]	**149.74** (104.49-258.63) [117.93; 73.01]
p-value	1.00		0.927	
Box-plots				

Table 6-6: ^{18}F-Fluoride PET Per-Patient Analysis; %Δ Volumetric Parameters; Response Analysis; (Mann-Whitney U tests for comparison)

6.2.3 %Δ Heterogeneity Parameters

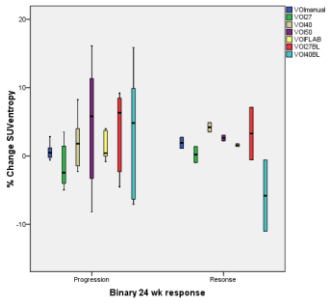
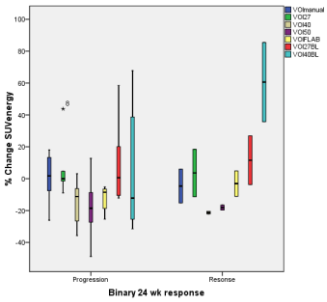
	%ΔSUV$_{entropy}$		%ΔSUV$_{energy}$	
	Progression	**Response**	**Progression**	**Response**
	Mean (Min-Max) [Median; SD] n=			
VOI$_{Manual}$ n=11	**-1.11** (-11.85-2.83) [0.19; 4.87] 7	**-0.78** (-3.6-2.72) [-1.13; 3.17] 4	**8.67** (-26.03-59.48) [4.46; 26.66] 7	**-13.69** (-23.57-5.95) [-18.58; 13.59] 4
p-value	0.927		0.230	
VOI$_{27}$ n=11	**-1.35** (-4.96-3.48) [-2.17; 3.02] 7	**-2.06** (-11.42-2.77) [0.2; 6.43] 4	**11.93** (-8.89-45.57) [0.69; 22.75] 7	**-22.05** (-59.83-18.44) [-23.41; 33.49] 4
p-value	1.00		0.109	
VOI$_{40}$ n=10	**2.01** (-2.29-8.24) [1.79; 3.83] 6	**-2.18** (-14.93-4.86) [3.53; 11.06] 3	**-14.63** (-35.71-3.12) [-11.2; 14.13] 6	**-16.86** (-22.1--8) [-20.49; 7.72] 3
p-value	1.00		0.905	
VOI$_{50}$ n=9	**4.59** (-8.21-16.13) [5.79; 9.05] 6	**-2.31** (-12.18-3.04) [2.21; 8.56] 3	**-18.19** (-48.76-12.71) [-18.57; 20.41] 6	**-21.63** (-28.63--16.57) [-19.68; 6.26] 3
p-value	0.262		0.714	
VOI$_{FLAB}$ N11	**1.06** (-0.83-3.99) [0.37; 1.96] 7	**0.77** (-0.4-1.77) [0.85; 0.99] 4	**-11.72** (-25.37--5.2) [-8.85; 7.44] 7	**-26.9** (-63.73-4.84) [-24.35; 30.17] 4
p-value	1.00		0.412	
VOI$_{27BL}$ N9	**3.91** (-4.54-9.19) [6.32; 5.83] 6	**-2.01** (-12.59-7.13) [-0.57; 9.94] 3	**9.51** (-12.21-58.48) [0.6; 27.24] 6	**-2.44** (-30.5-26.88) [-3.71; 28.71] 3
p-value	0.548		0.905	
VOI$_{40BL}$ n=8	**3.67** (-7.07-15.88) [4.82; 9.01] 6	**-5.82** (-11.04--0.6) [-5.82; 7.38] 2	**4.19** (-31.52-67.74) [-12.19; 40.12] 6	**60.6** (35.75-85.46) [60.6; 35.15] 2
p-value	0.286		0.286	
VOI$_{50BL}$ n=6	**-0.71** (-14.15-11.38) [-1.24; 9.32] 5	**n=1** **1.74**	**18.94** (-29.01-81.0) [12.68; 45.15] **5**	**n=1** **4.37**
p-value	n/a		n/a	
Box-plots				

Table 6-7: ^{18}F-Fluoride PET Per-Patient Analysis; %Δ SUV Heterogeneity Parameters; Response Analysis; (Mann-Whitney U tests for comparison)

The response analysis data can be reviewed in Tbl.6-7. The application of increasingly severe SUV thresholds for VOI definition resulted in a reduction in the number of patients with suitable

VOIs for quantification. A wider range of responses are seen when measuring the $\%\Delta SUV_{entropy}$ in patients who subsequently had PD confirmed, and the $\%\Delta SUV_{entropy}$ is higher this group. However, statistical comparison does not identify any significant difference between the distributions, but such analysis is limited by the small sample size.

There is no demonstrable difference identified in the distribution of $\%\Delta SUV_{energy}$ between the response groups. The per-lesion analysis a trend for a response in breast cancer lesion predicted with a change in SUV voxel distribution towards homogeneity (decrease entropy and increased energy).

No VOI method has demonstrated predictive superiority.

6.2.4 %ΔKi

There is no demonstrable difference in the distribution of %ΔKi between the response groups for these patients, irrespective of the VOI method used for quantification (see Tbl-6-8. The per-lesion analysis also did not demonstrate response predictive utility. Tumour sub-group response analyses are not feasible due to the small number of patients.

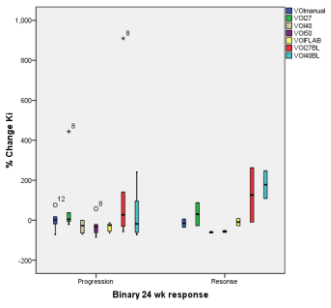
	%ΔKi	
	Progression	Response
	Mean (Min-Max) [Median; SD]	**Mean** (Min-Max) [Median; SD]
VOI_{Manual} n=11	**73.61** (-71.45-508.97) [10.45; 196.98] 7	**-27.16** (-45.73-5.99) [-34.46; 22.73] 4
p-value	0.164	
VOI_{27} n=11	**96.78** (-22.2-443.97) [11.61; 172.78] 7	**-17.76** (-88.55-87.86) [-35.18; 75.06] 4
p-value	0.109	
VOI_{40} n=9	**-31.18** (-69.72-0.01) [-26.47; 30.2] 6	**-30.56** (-63.24-28.51) [-56.95; 51.26] 3
p-value	1.00	
VOI_{50} n=9	**-29.28** (-83.8-57.86) [-32.95; 48.48] 6	**-52.63** (-60.86--45.47) [-51.57; 7.75] 3
p-value	0.548	
VOI_{FLAB} n11	**-31.78** (-64.16--12.98) [-24.11; 19.58] 7	**-42.81** (-90.83-8.42) [-44.41; 43] 4
p-value	0.648	
VOI_{27BL} N9	**169.48** (-56.7-909.73) [27.2; 370.67] 6	**74.68** (-28.65-262.35) [-9.67; 162.81] 3
p-value	0.714	
VOI_{40BL} n=8	**28.2** (-71.68-241.96) [-17.9; 120.55] 6	**177.6** (108.52-246.68) [177.6; 97.69] 2
p-value	0.143	
VOI_{50BL} n=6	**112.97** (-60.18-513.25) [-6.34; 240.49] 5	**-4.55** n=1
p-value	n/a	
Box-plots		

Table 6-8: ^{18}F-Fluoride PET Per-Patient Analysis; %Δ Ki; Response Analysis; (Mann-Whitney U tests for comparison)

6.3 Discussion

No utility has been demonstrated in this per-patient analysis of the ^{18}F-Fluoride scan for predicting a treatment response. However, with the small sample, lack of utility also cannot be concluded. The per-lesion analysis showed predictive utility from the behaviour of ^{18}F-Fluoride tracer uptake by bone metastases. The main limitation with this study is the size of the sample group. The per-lesion analysis clearly demonstrated differences between the parameter characteristics between breast and prostate cancer bone metastases, and analysis of all these lesions together may hide clinical associations. It would be ideal to identify imaging parameters with consistency between tumour subtypes, but the ^{18}F-Fluoride uptake is, by design, affected by osteoblastic activity and therefore will be different for tumours depending on what peri-tumour bone reaction is predominant.

The VOI methods were shown in the previous chapter to be result in statistically significant differences of the calculated quantification parameters. It has not been possible to demonstrate superiority of a VOI method given the negative nature of this analysis.

Chapter 7 ^{18}F Fluoride PET Per-Patient Analysis – Survival Analysis

The methods of this chapter and the statistical approaches for this section can be reviewed in Chapter 2.3.6.

7.1 OS Analysis

7.1.1 Tumour Group Analysis

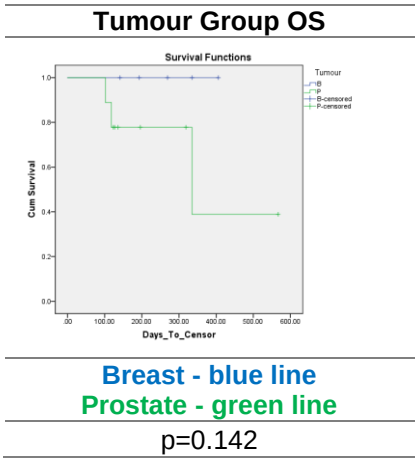


Figure 7-1: ^{18}F-Fluoride PET Per Patient Analysis; KM OS Analysis; Tumour Subgroups; B=Breast cancer P=Prostate Cancer

The breast cancer patients have an OS benefit, although significance has not been reached (p=0.142, see Fig.7-1); this benefit is real, but the small sample size has impaired statistical analysis.

OS analysis has therefore not been reported at the per-patient level in this thesis due to the small number of patients (14) and deaths (3).

7.2 PFS

7.2.1 Tumour Group Analysis

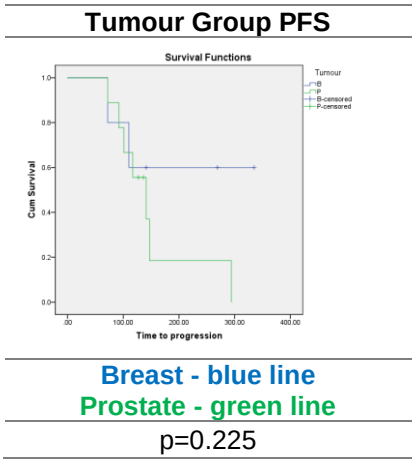


Figure 7-2: ^{18}F-Fluoride PET; KM PFS Analysis; Tumour Subgroups; B=Breast cancer P=Prostate Cancer

The per-lesion analysis demonstrated the significant PFS benefit of the breast cancer patients. Significance is not met with per-patient analysis due to the small sample group (Fig.7-2).

7.2.2 Baseline Scan

13 patients (5 Breast (2 with PD), 8 Prostate (7 with PD)) had information available for PFS analysis. Subgroup analysis of the breast patients is therefore not feasible.

7.2.2.1 SUV Parameters

The results of the Cox analysis are shown in the Table 7-1:

	SUV_{mean}	SUV_{max}	SUV_{peak}
	Cox regression p-value [HR]		
VOI_{manual}	0.606 [1.033]	0.633 [1.005]	0.630 [1.007]
VOI_{27}	0.57 [1.018]	0.657 [1.005]	0.658 [1.006]
VOI_{40}	0.609 [1.014]	0.657 [1.005]	0.647 [1.007]
VOI_{50}	0.628 [1.012]	0.657 [1.005]	0.638 [1.006]
VOI_{FLAB}	0.582 [1.018]	0.653 [1.005]	0.653 [1.006]

Table 7-1: ^{18}F-Fluoride PET Per-Patient Analysis; Baseline SUV Parameters; Cox regression univariate PFS Analysis; All Patients; all threshold methods.

No significant correlation has been identified in these patients between the SUV parameters analysed and risk of PD. Table 7-2 reports KM analysis; there is no statistically significant separation of the PFS curves for any SUV parameter.

Analysis of the prostate cancer lesions also demonstrates no statistically significant association between the SUV parameters, but suggests patients with a higher baseline SUV have a trend towards a longer PFS (Tbl.7-3).

The per-lesion analysis did not reveal any association between the baseline SUV parameters and PFS for these patients.

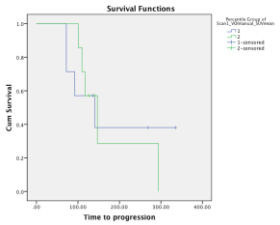
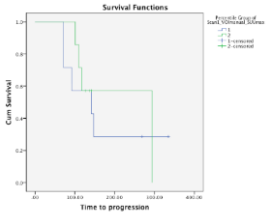
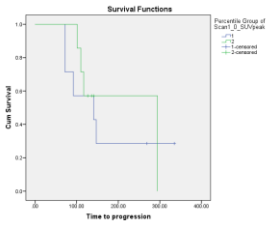
	$SUV_{mean}^{VOImanual}$	$SUV_{max}^{VOImanual}$	$SUV_{peak}^{VOImanual}$
		Median PFS in days (95% CI)	
			
Median	6.57	34.39	27.12
≤ (1)	141 (40.49-241.51)	141 (15.25-266.75)	141 (15.25-266.75)
> (2)	147 (101.18-192.82)	Not reached	Not reached
p-Value	0.933	0.751	0.751

Table 7-2: ^{18}F-Fluoride PET Per-Patient Analysis; Baseline SUV Parameters; KM PFS Analysis; All Patients; all threshold methods.

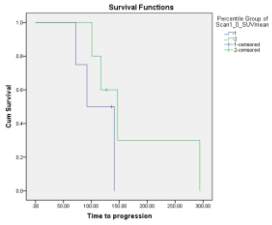
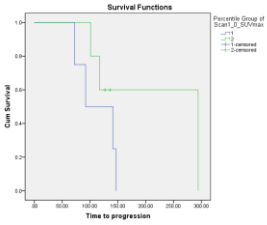
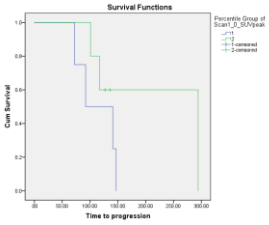
	$SUV_{mean}^{VOImanual}$ **Prostate patients only**	$SUV_{max}^{VOImanual}$ **Prostate patients only**	$SUV_{peak}^{VOImanual}$ **Prostate patients only**
			
		Low (1) - blue line High (2) - green line	
p-Value	0.248	0.235	0.235

Table 7-3: ^{18}F-Fluoride PET Per-Patient Analysis; Baseline SUV Parameters; KM PFS Analysis; Prostate Patients only; all threshold methods.

7.2.2.2 Volumetric Parameters

	MTV	**TLA**
	Cox regression p-value [HR]	
VOI_{manual}	0.272 [1.001]	0.212 [1.000]
VOI_{27}	0.242 [1.003]	0.167 [1.000]
VOI_{40}	0.211 [1.005]	0.154 [1.000]
VOI_{50}	0.202 [1.007]	0.152 [1.000]
VOI_{FLAB}	0.218 [1.003]	0.179 [1.000]

Table 7-4: ^{18}F-Fluoride PET Per-Patient Analysis; Baseline Volumetric Parameters; Cox regression univariate PFS Analysis; All Patients; all threshold methods.

Cox regression analysis (Tbl.7-4) does not demonstrate a correlation between PFS and the volumetric parameters. There is trend towards a significant separation of the KM survival curve for MTV (Tbl.7-5). However, breast cancer patients had smaller lesions and a longer PFS. KM analysis of the prostate patients only reveals no PFS prediction utility (p=0.629), and TMV/TLA are not independent predictors.

The per-lesion analysis suggested a PFS benefit for the breast cancer lesions with a lower MTV at baseline, but there are too few breast cancer patients sub-group analysis.

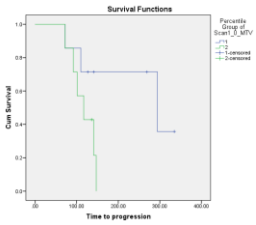
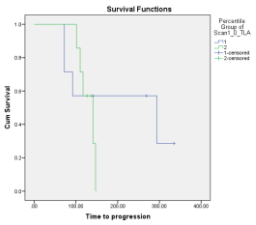
	$MTV^{VOImanual}$	$TLA^{VOImanual}$
	Median PFS in days (95% CI)	
		
Median	326.82	2520.3
≤ Median (1)	294 (24.81-563.19)	294 (0-602.50)
> Median (2)	117 (75.94-158.06)	141 (104.35-177.65)
p-Value (Log Rank)	0.055	0.442

Table 7-5: ^{18}F-Fluoride PET Per-Patient Analysis; Baseline Volumetric Parameters; KM PFS Analysis; All Patients; all threshold methods.

7.2.2.3 Heterogeneity Parameters

	$SUV_{entropy}$	SUV_{energy}
	Cox regression p-value [HR]	
VOI_{manual}	0.691 [1.346]	0.718 (0.035)
VOI_{27}	0.713 [1.242]	0.641 (0.011)
VOI_{40}	0.779 [1.180]	0.708 (0.037)
VOI_{50}	0.560 [1.453]	0.600 (0.006)
VOI_{FLAB}	0.667 [1.321]	0.747 (0.061)

Table 7-6: ^{18}F-Fluoride PET Per-Patient Analysis; Baseline SUV Heterogeneity Parameters; Cox regression univariate PFS Analysis; All Patients; all threshold methods.

For these patients, there is no demonstrable correlation between the baseline heterogeneity parameters and PFS using univariate Cox-Regression (Tbl.7-6). There is no significant separation of the KM survival curves (Tbl.7-7). No association with PFS was identified with per-lesion analysis either.

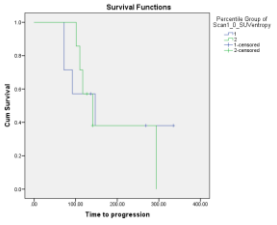
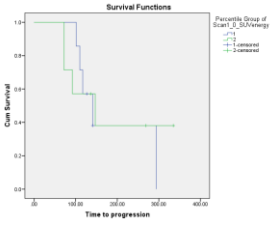
	$SUV_{entropy}^{VOImanual}$	$SUV_{energy}^{VOImanual}$
	Median PFS in days (95% CI)	
		
Median	2.67	0.09
≤ Median (1)	147 (34.18-259.82)	141 (91.77-190.23)
> Median (2)	141 (91.77-190.23)	147 (34.18-259.82)
p-Value (Log Rank)	0.796	0.796

Table 7-7: ^{18}F-Fluoride PET Per-Patient Analysis; Baseline SUV Heterogeneity Parameters; KM PFS Analysis; All Patients; all threshold methods.

7.2.2.4 Uptake Rate Constant (Ki)

	Ki
	Cox regression p-value [HR]
VOI_{manual}	0.833 [2.432]
VOI_{27}	0.768 [1.871]
VOI_{40}	0.803 [1.579]
VOI_{50}	0.656 [2.116]
VOI_{FLAB}	0.787 [1.822]

Table 7-8: ^{18}F-Fluoride PET Per-Patient Analysis; Baseline Ki; Cox regression univariate PFS Analysis; All Patients; all threshold methods.

There is no demonstrable correlation between PFS and the Ki measured on the baseline scan (Tbl.7-8). KM analysis is in Table 7-9; there is no significant separation of the survival curves. Per lesion analysis also did not demonstrate an association between baseline Ki and PFS.

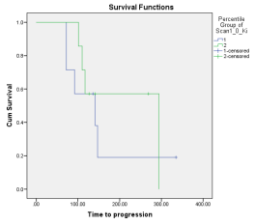
	Ki
	Median PFS in days (95% CI
	
Median	0.09
≤ Median (1)	141 (40.49-241.51)
> Median (2)	Not reached
p-Value (Log Rank)	0.553

Table 7-9: ^{18}F-Fluoride PET Per-Patient Analysis; Baseline Ki; KM PFS Analysis; All Patients; all threshold methods.

7.2.3 Percentage Change (Δ%) Between Scans - PFS Analysis

n=11 (5 Breast, 6 Prostate), 8 events (2 Breast, 6 Prostate)

7.2.3.1 Δ% SUV Parameters

Univariate Cox regression analysis for these patients can be reviewed in Table 7-10. No correlation has been identified in these patients between the time to progression and the size of Δ%SUV parameter.

	Δ%SUV$_{mean}$	**Δ%SUV$_{max}$**	**Δ%SUV$_{peak}$**
	Cox regression p-value [HR]		
VOI$_{manual}$	0.243 [0.986]	0.808 [0.997]	0.804 [1.004]
VOI$_{27}$	0.809 [1.003]	0.897 [1.002]	0.877 [1.002]
VOI$_{40}$	0.720 [1.005]	0.836 [1.002]	0.920 [1.001]
VOI$_{50}$	0.695 [1.005]	0.824 [1.002]	0.996 [1.000]
VOI$_{FLAB}$	0.737 [0.996]	0.938 [1.001]	0.814 [1.004]
VOI$_{27BL}$	0.626 [0.997]	0.362 [1.003]	0.378 [1.005]
VOI$_{40BL}$	0.627 [0.996]	0.300 [1.003]	0.323 [1.005]
VOI$_{50BL}$	0.656 [0.996]	0.244 [1.004]	0.249 [1.005]

Table 7-10: ^{18}F-Fluoride PET Per-Patient Analysis; %Δ SUV Parameters; Cox regression univariate PFS Analysis; All Patients; all threshold methods.

KM analysis of the VOI$_{manual}$ data can be reviewed in Table 7-11. The plots are similar for all 3 parameters, showing a statistical trend for PFS benefit for patients with a more positive change in the SUV. %Δ SUV$_{max}$ is perhaps the more discriminating parameter, demonstrating a 177-day median-PFS benefit for patients with %ΔSUV$_{max}$>1.13% (p=0.056).

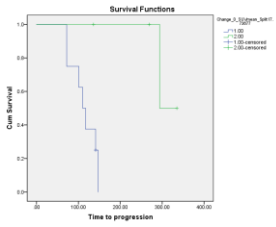
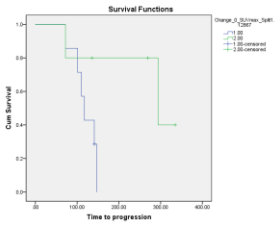
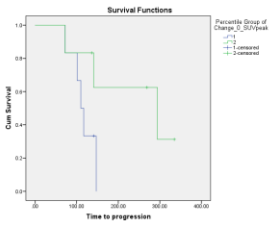
	Δ%SUV$_{mean}^{VOImanual}$	**Δ%SUV$_{max}^{VOImanual}$**	**Δ%SUV$_{peak}^{VOImanual}$**
	Median PFS in days (95% CI)		
			
Dichotomisation	+17.74%	1.1286	-1.99% (median)
≤ Dichotomisation	110 (87.83-132.17)	117 (99.04-134.96)	110 (91-129)
> Dichotomisation	294 (n/a)	294 (0-616.69)	294 (59-529)
p-Value (Log Rank)	0.09	0.056	0.097

Table 7-11: ^{18}F-Fluoride PET Per-Patient Analysis; %Δ SUV Parameters; KM PFS Analysis; All Patients; all threshold methods.

However, analysis of only the prostate cancer lesions reveals no association between the %Δ in SUV parameters and PFS (%Δ SUV$_{mean/max/peak}$, p=0.607). There are too few breast cancer patients for analysis. The prostate cancer patients had a shorter PFS compared with the breast cancer patients; the %Δ SUV parameters are not independent predictors.

7.2.3.2 Δ% Volumetric Parameters

Univariate Cox regression analysis can be reviewed in Table 7-12. No significant correlation has been identified in these patients.

	Δ%MTV	Δ%TLA
	Cox regression p-value [HR]	
VOI_{manual}	0.637 [1.002]	0.386 [0.998]
VOI_{27}	0.862 [0.999]	0.391 [1.015]
VOI_{40}	0.802 [0.999]	0.153 [1.027]
VOI_{50}	0.819 [0.999]	0.083 [1.031]
VOI_{FLAB}	0.843 [1.001]	0.653 [0.995]
VOI_{27BL}	0.821 [1.003]	0.751 [0.999]
VOI_{40BL}	0.939 [1.001]	0.946 [1.000]
VOI_{50BL}	0.984 [1.000]	0.989 [1.000]

Table 7-12: ^{18}F-Fluoride PET Per-Patient Analysis; %Δ Volumetric Parameters; Cox regression univariate PFS Analysis; All Patients; all threshold methods.

KM analysis (Tbl.7-13) demonstrates a statistically significant PFS advantage for patients with an increase in TLA between the scans, with a 193-day median-PFS advantage noted in the patients with greater than 5.8% increase between the scans (p=0.011). The change in MTV does not demonstrate a significant separation of the survival curves (p=0.296). TLA is the product of the MTV and SUV_{mean} – an increase in SUV_{mean} confers PFS advantage, and this effect is being translated through to %ΔTLA. In these patients, it seems %ΔTLA is more discriminating than %Δ SUV_{mean}.

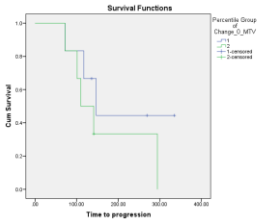
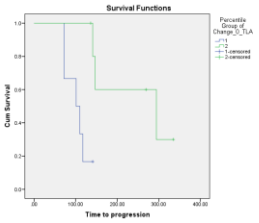
	Δ%MTV^{VOImanual}	**Δ%TLA^{VOImanual}**
	Median PFS in days (95% CI)	
		
Median	-0.35%	+5.82%
≤ Median (1)	147 (88.2-205.8)	101 (55.4-146.6)
> Median (2)	110 (62.0-158.0)	294 (64.7-523.3)
p-Value (Log Rank)	0.296	0.011

Table 7-13: ^{18}F-Fluoride PET Per-Patient Analysis; %Δ Volumetric Parameters; KM PFS Analysis; All Patients; all threshold methods.

Analysis of only the prostate cancer patients demonstrates no significant separation of the KM curve for %ΔMTV (p=0.951), but the survival curve separation with %ΔTLA is preserved and significant (Tbl.7-14).

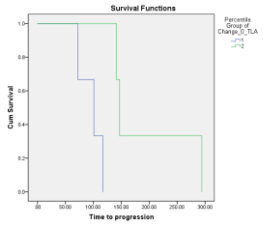
	Δ%TLA^{VOImanual}
	Median PFS in days (95% CI)
	
Median	-3.3%
≤ Median (1)	101 (54.6-147.4)
> Median (2)	147 (137.4-156.6)
p-Value (Log Rank)	0.025

Table 7-14: ^{18}F-Fluoride PET Per-Patient Analysis; %Δ SUV Parameters; KM PFS Analysis; Prostate Patients only; all threshold methods.

A 46-day median PFS benefit is predicted for prostate cancer patients with a decrease greater than 3% in the TLA between scan (p=0.025); this was shows for breast cancer lesions at per-lesion analysis.

7.2.3.3 Δ% SUV Heterogeneity Parameters

	Δ%SUV$_{entropy}$	**Δ%SUV$_{energy}$**
	Cox regression p-value [HR]	
VOI$_{manual}$	0.492 [0.925]	0.133 [1.029]
VOI$_{27}$	0.728 [0.976]	0.095 [1.025]
VOI$_{40}$	0.449 [1.060]	0.711 [0.989]
VOI$_{50}$	0.237 [1.069]	0.767 [0.993]
VOI$_{FLAB}$	0.857 [1.040]	0.359 [1.017]
VOI$_{27BL}$	0.246 [1.075]	0.896 [0.998]
VOI$_{40BL}$	0.495 [1.032]	0.362 [0.990]
VOI$_{50BL}$	0.235 [0.905]	0.211 [1.022]

Table 7-15: ^{18}F-Fluoride PET Per-Patient Analysis; %Δ SUV Heterogeneity Parameters; Cox regression univariate PFS Analysis; All Patients; all threshold methods.

Univariate Cox regression analysis of the parameters (Tbl.7-15), does not demonstrate any statistically significant correlation.

Table 7-16 reports the outcome of KM PFS analysis for these data. There is no statistically significant separation of the survival curves for either %Δ SUV parameter. However, there is the suggestion of a more negative change (less than +12.86%) being associated with a prolonged PFS (177 days for these patients, p=0.195). This is a shift towards the SUV voxel distribution having less uniformity (towards heterogeneity).

Analysis of the prostate cancer patients does not suggest an association between the %Δ heterogeneity parameters and PFS (log rank analysis: %ΔSUV$_{entropy/energy}$ p=0.277 and 0.988 respectively).

The per-lesion analysis demonstrated no confirmed association between these parameters and PFS, but suggests a trend for a PFS benefit for the breast cancer lesions with a decrease in $SUV_{entropy}$ and increase in SUV_{energy} between scans.

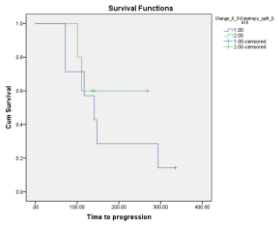
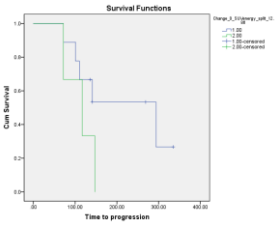
	$\Delta\%SUV_{entropy}^{VOImanual}$	**$\Delta\%SUV_{energy}^{VOImanual}$**
	Median PFS in days (95% CI)	
KM Plots		
Dichotomisation	0.915	12.86
≤ Dichotomisation (1)	141 (79.41-202.59)	294 (106.91-481.09)
> Dichotomisation (2)	Not reached	117 (44.99-189.01)
p-Value (Log Rank)	0.558	0.195

Table 7-16: ^{18}F-Fluoride PET Per-Patient Analysis; %Δ SUV Heterogeneity Parameters; KM PFS Analysis; All Patients; all threshold methods.

7.2.3.4 Δ% Ki

	Δ%Ki
	Cox regression p-value [HR]
VOI_{manual}	0.088 [1.005]
VOI_{27}	0.289 [1.002]
VOI_{40}	0.603 [0.995]
VOI_{50}	0.879 [1.001]
VOI_{FLAB}	0.975 [1.000]
VOI_{27BL}	0.871 [1.000]
VOI_{40BL}	0.363 [0.997]
VOI_{50BL}	0.139 [1.010]

Table 7-17: ^{18}F-Fluoride PET Per-Patient Analysis; %Δ Ki; Cox regression univariate PFS Analysis; All Patients; all threshold methods.

The Cox regression analysis (Tbl.7-17) does not demonstrate any significant correlation.

The KM analysis (Tbl.7-18) does not demonstrate a statistically significant separation of the survival curves, but there is perhaps the suggestion of a PFS benefit from a decrease in Ki between scans (p=0.152).

However, analysis of the prostate cancer patients only does not demonstrate an association between %Δ Ki and PFS (p=0.988). There are too few breast patients for useful analysis, but it is likely Ki analysis in not independent to the tumour types.

Per-lesion analysis did not demonstrate any PFS prognostic utility for these patients.

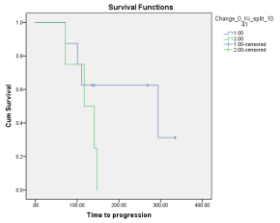
	Δ%Ki
	Median PFS in days (95% CI)
KM Plot	
Median	10.41
≤ Median (1)	294 (20.53-567.47)
> Median (2)	117 (49.38-184.62)
p-Value (Log Rank)	0.152

Table 7-18: ^{18}F-Fluoride PET Per-Patient Analysis; %Δ Ki; KM PFS Analysis; All Patients; all threshold methods.

7.3 Discussion

The per-lesion analysis of these patients showed prognostic utility of the quantification parameters; this has not been elicited with per-patient analysis.

The %Δ parameters at the per-lesion analysis suggested a PFS benefit for breast cancer patients with a decrease in TLA between scans, a fall in $SUV_{entropy}$ and an increase in SUV_{energy}. The breast cancer patients cannot be analysed at the per-patient level, but the prostate cancer patients have been shown to have a PFS advantage from an increase in TLA between the scans (p=0.025) which is contradictory to what analysis the breast cancer lesions suggested. The TLA is a product of the MTV and SUV_{mean}, and the change in neither of these parameters have shown prognostic utility. The %ΔTLA of the per-patient analysis has not shown an association with treatment response. An increase in TLA will be caused by an increase in MTV or an increase in SUV_{mean}. The MTV for the ^{18}F-Fluoride PET is not a direct measure of the tumour deposit size, but is a volume of bone with an osteoblastic reaction due to the interaction of the tumour cells; a flare response of a healing bone metastasis might cause both an increased tracer uptake (therefore SUV_{mean}) and may show an increased volume of reactive bone, and hence MTV. This needs to be examined further with a larger patient population, including both breast and prostate cancer patients to examine the differences between the tumour types.

The per-patient survival analysis has therefore been disappointing, the likelihood of prognostication from ^{18}F-Fluoride PET having been suggested with the per-lesion analysis. This study is likely to be underpowered to identify differences if they are smaller, even if real clinically relevant information does exist. The multiple analyses should be statistically corrected for due to the increased risk of identifying an association through chance alone; this correction has not been applied to this analysis. However, this analysis provides data for generating new hypothesis, and for taking forward methods into subsequent prospective trials for further evaluation.

Chapter 8 ^{18}F-Fluoride PET Whole-Body Quantification - Appraisal of novel whole-skeletal ^{18}F-Fluoride PET quantification parameters for predictive and prognostic utility in the assessment of bone metastases from breast and prostate cancer.

The methods of this chapter and the statistical approaches for this section can be reviewed in Chapter 2.3.7. This chapter is presented in 4 sections: 8.1-Descriptive statistics and response analysis; 8.2-OS Analysis; 8.3-PFS Analysis; 8.4-Discussion.

8.1 Response Analysis

8.1.1 Baseline Scan

14 patients had baseline ^{18}F-Fluoride PET imaging (BCa-5, PCa-9). 12 patients had data for response analysis (BCa-5, PCa-7). 4 patients had a treatment response (including stable disease) maintained at 24-weeks (BCa-3,PCa-1)

8.1.1.1 SUV Parameters

The whole-population distributions (Tbl.8-1) of the SUV parameters are significantly different from each other (Friedman's Variance test, $p<0.0001$). The distributions medians of SUV_{max} and SUV_{peak} are higher than SUV_{mean}; SUV_{mean} will involve more lower value voxels. No statistical difference in the parameter distributions is identified between tumour groups, but the distribution mean for the prostate patients is higher than the breast patients; there are too few patients to statistically confirm this.

There is no demonstrable difference in the distributions of the SUV parameters between the response groups (Mann-Whitney U tests: $SUV_{mean/max/peak}$ p=1.0, 0.933, 0.93 respectively).

The individual patient response data for each parameter split can be seen in the waterfall plots (Figs.8-1,8-2,8-3). The highest SUV_{WB} measurements were from patients with PD, but the SUV_{WB} was lower in some PD patients compared to those a response; identifying a discriminating threshold is not possible.

Response analysis of only the breast cancer patients does not demonstrate an significant difference in the distributions between the response groups (p=1.0 for all parameters), but with so few patients this does not exclude a difference.

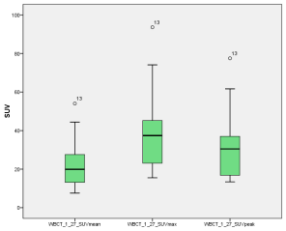
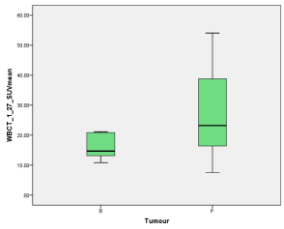
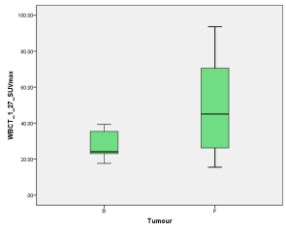
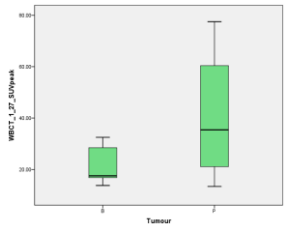
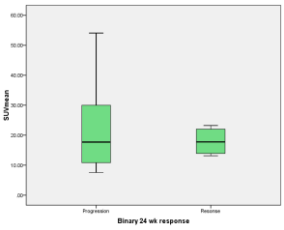
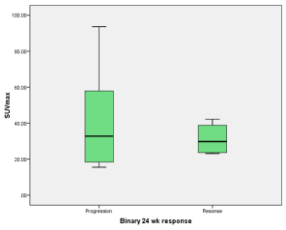
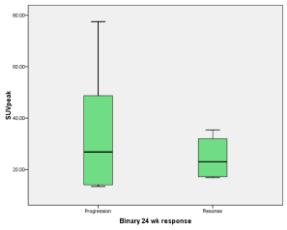
SUV$_{meanWB}$	**SUV$_{maxWB}$**	**SUV$_{peakWB}$**
Mean (Min-Max) [Median; SD]		
23.04 (7.56-54.05) [19.92; 13.76]	**40.82** (15.53-93.64) [37.41; 23.71]	**33.19** (13.39-77.52) [30.48; 20.28]
		
Breast **n=5**	**Prostate** **n=9**	**Breast** n=5	**Prostate** **n=9**	**Breast** n=5	**Prostate** **n=9**
16.13 (10.86-21.15) [14.68; 4.64]	**26.87** (7.56-54.05) [23.22; 15.83]	**27.95** (17.69-39.37) [24.13; 9.09]	**47.97** (15.53-93.64) [45.1; 26.67]	**21.81** (13.75-32.51) [17.53; 8.17]	**39.51** (13.39-77.52) [35.4; 22.56]
					
p=0.298		p=0.147		p=0.147	
Progression n=8 (BCa 2; PCa 6)	**Response** n=4 (BCa 3; PCa 1)	**Progression** **n=8** (BCa 2; PCa 6)	**Response** n=4 (BCa 3; PCa 1)	**Progression** **n=8** (BCa 2; PCa 6)	**Response** n=4 (BCa 3; PCa 1)
22.33 (7.56-54.05) [17.71; 16.07]	**17.97** (13.14-23.22) [17.76; 4.83]	**40.92** (15.53-93.64) [32.81; 28.18]	**31.22** (23.1-42.21) [29.79; 9.22]	**33.75** (13.39-77.52) [26.8; 23.87]	**24.55** (16.82-35.4) [22.99; 8.98]
					
p=1.00		p=0.933		0.933	

Table 8-1: ^{18}F-Fluoride PET; Whole-Body Analysis; Baseline Scan SUV parameters; all patients; Descriptive statistics and response analyses; Mann-Whitney U tests for comparisons between distributions (B=Breast, P=Prostate)

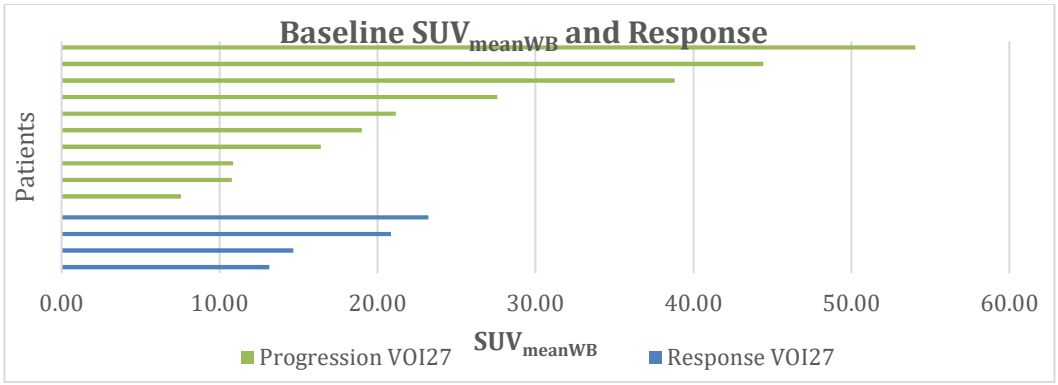


Figure 8-1: ^{18}F-Fluoride PET; Whole-Body Analysis; Baseline SUV_{meanWB}^{VOI27}; all patients split into response groups (Progression=Green; Response=Blue)

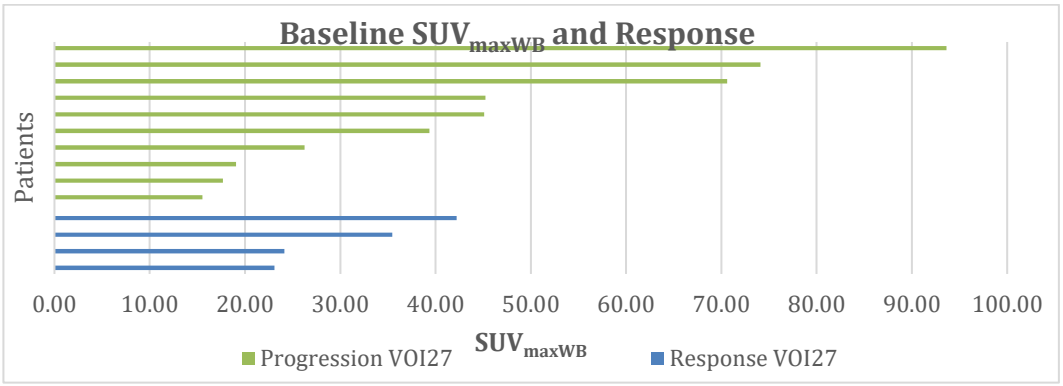


Figure 8-2: ^{18}F-Fluoride PET; Whole-Body Analysis; Baseline SUV_{maxWB}^{VOI27}; all patients split into response groups (Progression=Green; Response=Blue)

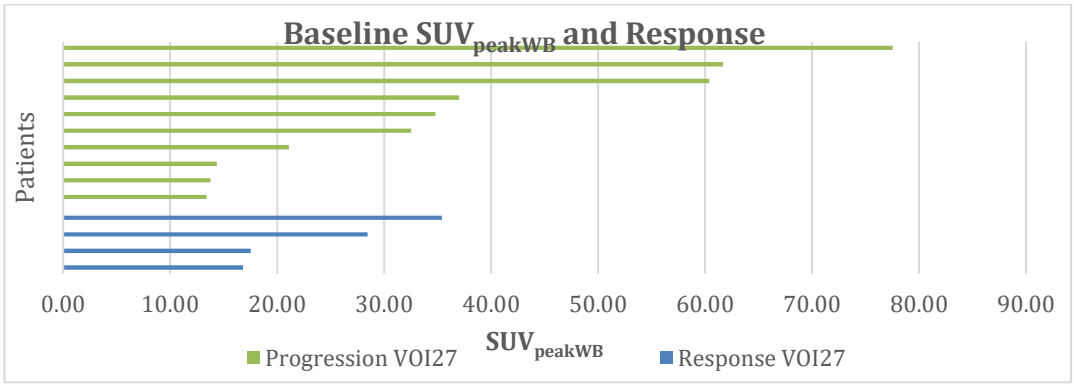


Figure 8-3: ^{18}F-Fluoride PET; Whole-Body Analysis; Baseline SUV_{peakWB}^{VOI27}; all patients split into response groups (Progression=Green; Response=Blue)

8.1.1.2 Volumetric Parameters

Table 8-2 reports the data; there is no statistical difference in the MTV_{WB} between the tumour groups (p=0.298), but there is a trend suggesting larger MTVs for prostate patients. TLA_{WB} is significantly larger for the prostate patients (p=0.042); combining the higher baseline SUV_{meanWB} and higher MTV_{WB} for the prostate lesions, neither of which were discriminating in themselves, to create a more discriminating parameter.

The mean MTV_{WB} is greater for those with PD, but this is not significant (p=0.570). The largest MTV_{WB} values were identified in patients with PD. Similarly, the calculated TLA_{WB} is larger in those with PD, but this is not significant (p=0.368).

Figures 9-4 and 9-5 represent the individual patient data for MTV_{WB} and TLA_{WB} as waterfall plots.

It is not possible to identify values of MTV_{WB} or TLA_{WB} to accurately predict a treatment response in these patients. The prostate cancer patients had a larger MTV and TLA compared to the breast patients; these differences are significant for TLA (p=0.042) but not for MTV (p=0.298). The prostate patients were much less likely to have a treatment response.

Response analyses for just the breast cancer patients can be seen in Table 8-3. Statistical comparisons are compromised by the small number of patients, but the breast patients with PD had larger MTV and TLA, although this has not reached significance (p=0.4 for both parameters).

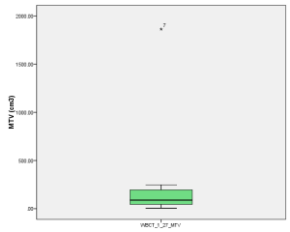
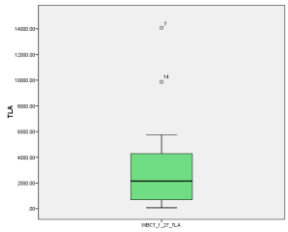
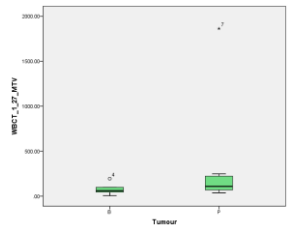
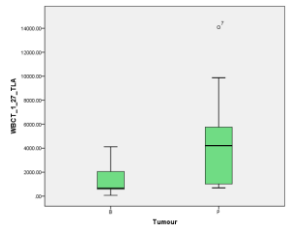
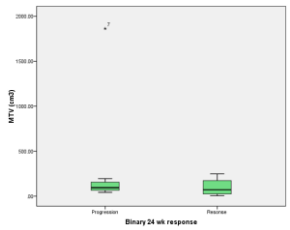
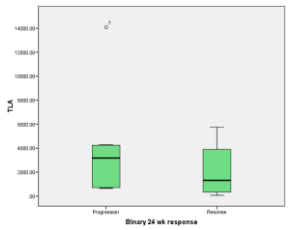
MTV_{WB}		**TLA_{WB}**	
Mean (Min-Max) [Median; SD]			
227.9 (4.86-1863.3) [88.87; 476.27]		**3599.64** (71.31-14082.86) [2148.67; 4045.24]	
			
Breast n=5	**Prostate** n=9	**Breast** n=5	**Prostate** n=9
80.6 (4.86-194.64) [60.65; 72.06]	**309.74** (37.02-1863.3) [108.56; 587.29]	**1496.3** (71.31-4115.95) [658.51; 1638.5]	**4768.16** (691.75-14082.86) [4212.65; 4576.9]
			
p=0.298		p=0.042	
Progression n=8 (BCa 2; PCa 6)	**Response** n=4 (BCa 3; PCa 1)	**Progression** **n=8** (BCa 2; PCa 6)	**Response** n=4 (BCa 3; PCa 1)
316.96 (42.15-1863.3) [93.91; 626.59]	**98.89** (4.86-247.87) [71.42; 106.47]	**3879.36** (658.51-14082.86) [3180.29; 4431.69]	**2115.78** (71.31-5756.05) [1317.87; 2568.05]
			
p=0.570		p=0.368	

Table 8-2: ^{18}F-Fluoride PET; Whole-Body Analysis; Baseline Scan Volumetric Parameters; all patients; Descriptive statistics and response analyses; Mann-Whitney U tests for comparisons between distributions (B=Breast, P=Prostate)

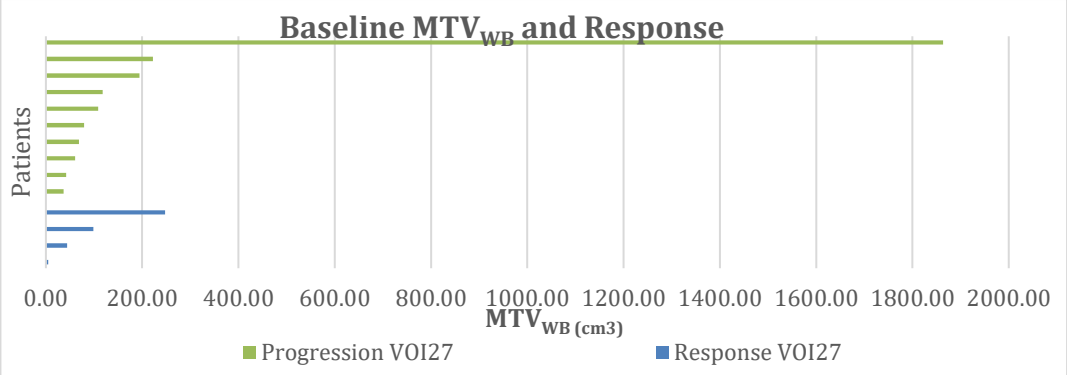


Figure 8-4: ^{18}F-Fluoride PET; Whole-Body Analysis; Baseline MTV$_{WB}^{VOI27}$; all patients split into response groups (Progression=Green; Response=Blue)

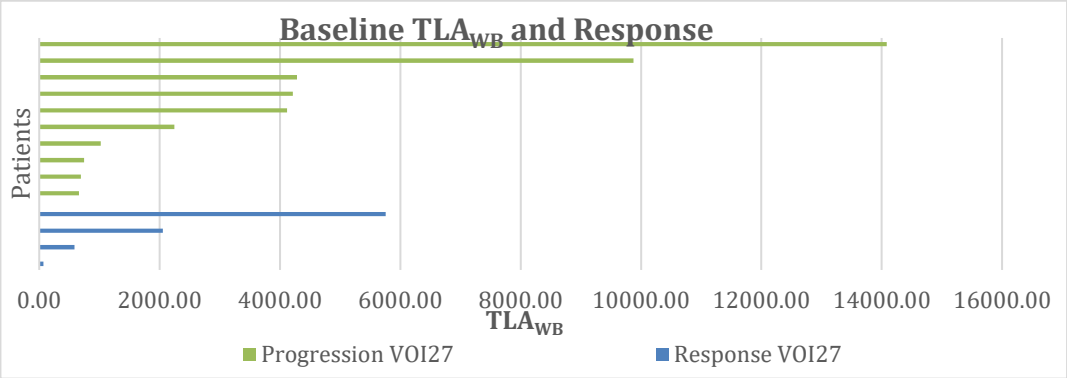


Figure 8-5: ^{18}F-Fluoride PET; Whole-Body Analysis; Baseline TLA$_{WB}^{VOI27}$; all patients split into response groups (Progression=Green; Response=Blue)

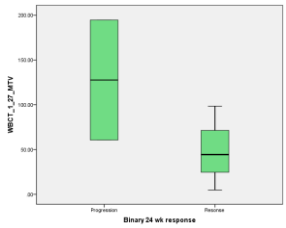
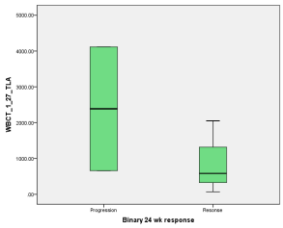
MTV$_{WB}$ **Breast only**		**TLA$_{WB}$** **Breast only**	
Mean (Min-Max) [Median; SD] n=5			
Progression n=2	**Response** n=3	**Progression** n=2	**Response** n=3
127.64 (60.65-194.64) [127.64; 94.74]	**49.23** (4.86-98.47) [44.37; 47]	**2387.23** (658.51-4115.95) [2387.23; 2444.77]	**902.35** (71.31-2052.7) [583.04; 1028.57]
			
p=0.4		p=0.4	

Table 8-3: ^{18}F-Fluoride PET; Whole-Body Analysis; Baseline Scan Volumetric Parameters; Breast Patients Only; Descriptive statistics and response analyses; Mann-Whitney U tests for comparisons between distributions

8.1.1.3 Heterogeneity Parameters

The data is recorded in Table 8-4. There is a wider range of whole-body SUV heterogeneity parameters in those patients with PD, but there were twice the number of patients in this group.

The prostate cancer patients had a higher $ADC_{entropy}$ and lower ADC_{energy} than the breast cancer patients; these differences do not reach statistical significance (p=0.190 and 0.083 respectively), but this suggests a trend. No statistically significant difference has been identified in the distributions of parameters between the response groups. Both the highest and lowest parameters were identified in patients with PD; it is not possible to identify a value to usefully predict treatment response. The waterfall plots (Figs.8-6,8-7) display the individual patient data for comparison.

The distribution of the breast cancer heterogeneity parameters are also not significantly different between the response groups (p=1.0 for each parameter), but with so few patients it is not possible to draw conclusions from this analysis.

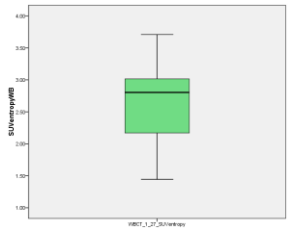
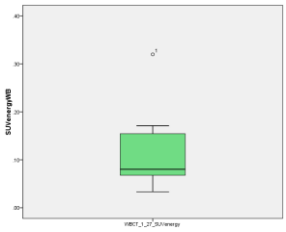
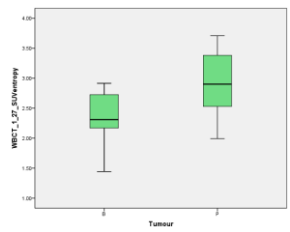
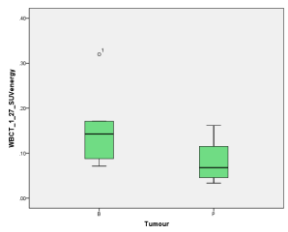
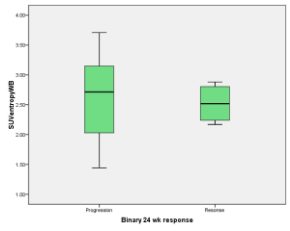
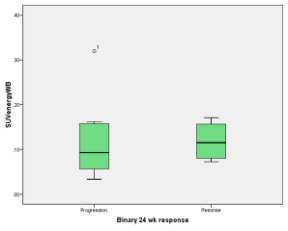
$SUV_{entropyWB}$		**$SUV_{energyWB}$**	
Mean (Min-Max) [Median; SD]			
2.69 (1.44-3.71) [2.8; 0.65]		**0.11** (0.03-0.32) [0.08; 0.08]	
			
Breast n=5	**Prostate** n=9	**Breast** n=5	**Prostate** n=9
2.31 (1.44-2.92) [2.31; 0.57]	**2.9** (1.99-3.71) [2.9; 0.62]	**0.1586** (0.0714-0.3199) [0.1428; 0.0988]	**0.0836** (0.0333-0.1616) [0.0679; 0.0489]
			
p=0.190		p=0.083	
Progression n=8 (BCa 2; PCa 6)	**Response** n=4 (BCa 3; PCa 1)	**Progression n=8** (BCa 2; PCa 6)	**Response** n=4 (BCa 3; PCa 1)
2.62 (1.44-3.71) [2.72; 0.76]	**2.52** (2.17-2.88) [2.52; 0.34]	**0.1212** (0.0337-0.3199) [0.0932; 0.0933]	**0.1186** (0.0728-0.171) [0.1153; 0.0461]
			
p=0.808		p=0.570	

Table 8-4: ^{18}F-Fluoride PET; Whole-Body Analysis; Baseline Scan SUV Heterogeneity Parameters; all patients; Descriptive statistics and response analyses; Mann-Whitney U tests for comparisons between distributions (B=Breast, P=Prostate)

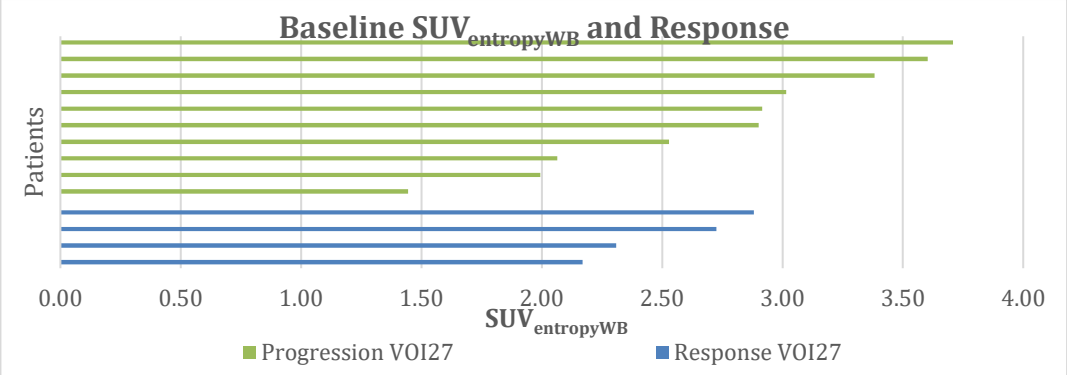


Figure 8-6: ^{18}F-Fluoride PET; Whole-Body Analysis; Baseline $SUV_{entropyWB}^{VOI27}$; all patients split into response groups (Progression=Green; Response=Blue)

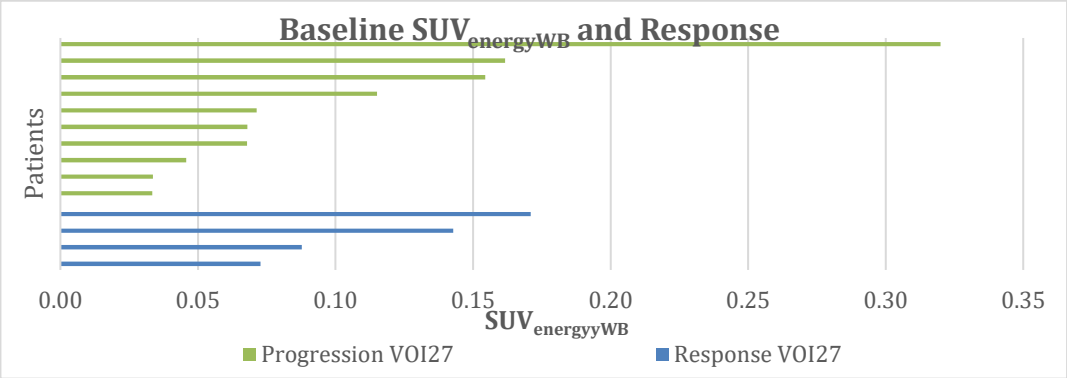


Figure 8-7: ^{18}F-Fluoride PET; Whole-Body Analysis; Baseline $SUV_{energyWB}^{VOI27}$; all patients split into response groups (Progression=Green; Response=Blue)

8.1.1.4 Ki_{WB}

11 patients had data suitable for analysis, 5 breast patients and 6 prostate patients. The data is displayed in Table 8-5.

There is no demonstrable difference in the Ki_{WB} between the tumour groups (p=1.0). Baseline Ki_{WB} does not differentiate between patients who subsequently are defined as responders and those with PD (p=0.788).

The waterfall plot (Fig.8-8) shows the highest Ki_{WB} values were recorded in patients with PD, but this response group also included patients with the lowest Ki_{WB} measurements from the baseline scan. No useful cut off to accurately separate patients into response groups cannot be identified. Breast cancer lesions have not been analysed separately because there is no demonstrable difference in the Ki_{WB} between the tumour groups.

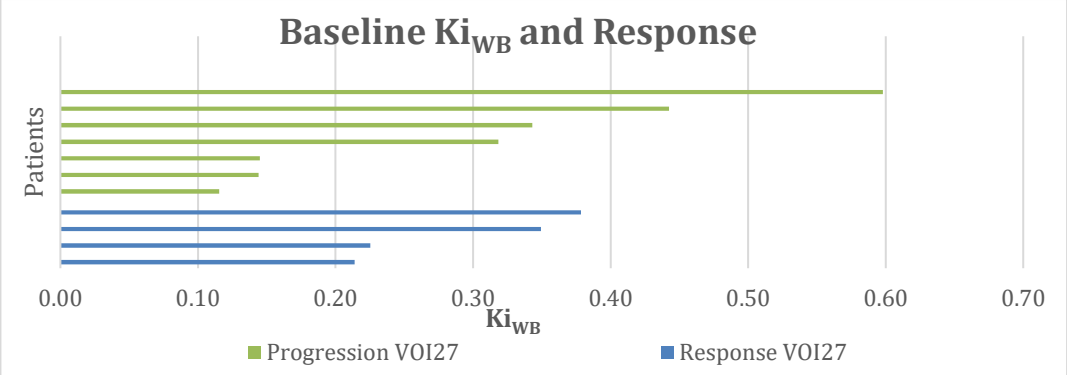


Figure 8-8: ^{18}F-Fluoride PET; Whole-Body Analysis; Baseline Ki$_{WB}^{VOI27}$; all patients split into response groups (Progression=Green; Response=Blue)

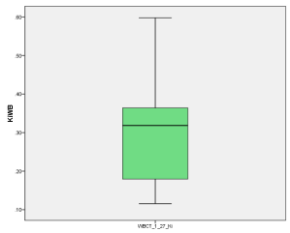
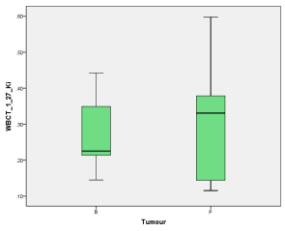
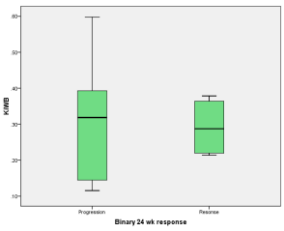
Ki$_{WB}^{VOI27}$ **Mean** (Min-Max) [Median; SD]				
All pts	**Breast** n=5	**Prostate** n=6	**Progression** n=7 (BCa 2; PCa 5)	**Response** n=4 (BCa 3; PCa 1)
0.24 (0.08-0.5) [0.25; 0.12]	**0.28** (0.15-0.44) [0.23; 0.12]	**0.32** (0.12-0.6) [0.33; 0.18]	**0.3** (0.12-0.6) [0.32; 0.18]	**0.29** (0.21-0.38) [0.29; 0.08]
				
	p=1.0		p=0.788	

Table 8-5: ^{18}F-Fluoride PET; Whole-Body Analysis; Baseline Scan Ki; all patients; Descriptive statistics and response analyses; Mann-Whitney U tests for comparisons between distributions (B=Breast, P=Prostate)

8.1.2 Second Scan

n=12 (5 breast cancer patients, 7 prostate cancer patients).

The table and box plots below summarise the SUV parameter data from the 2nd scan:

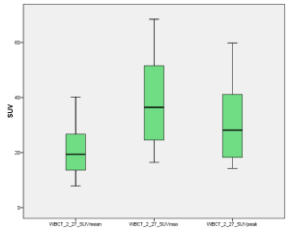
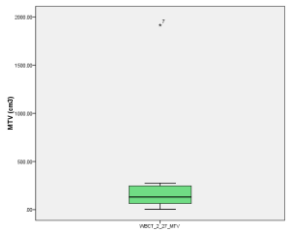
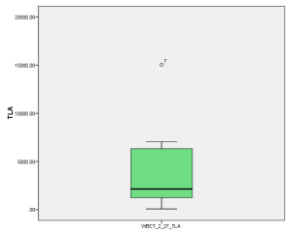
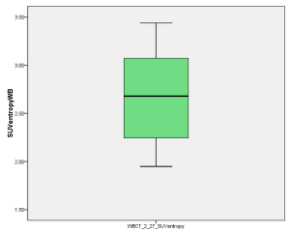
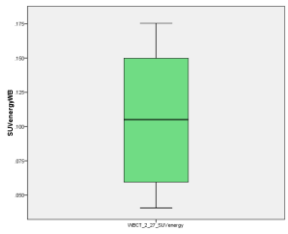
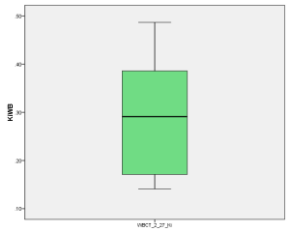
	Mean (Min-Max) [Median; SD]	**Box-Plot**
SUV_{meanWB}	**21.05** (7.85-40.22) [19.37; 10.11]	
SUV_{maxWB}	**38.52** (16.46-68.46) [36.44; 18.03]	
SUV_{peakWB}	**30.83** (14.27-59.79) [28.15; 15.28]	
MTV_{WB}^{VOI27} cm^3	**283.73** (3.87-1915.38) [131.32; 521.39]	
TLA_{WB}	**4135.23** (77.46-15044.88) [2143.78; 4270.98]	
$SUV_{entropyWB}$	**2.68** (1.95-3.44) [2.68; 0.51]	
$SUV_{energyWB}$	**0.1** (0.04-0.18) [0.11; 0.05]	
Ki_{WB} n=11	**0.22** (0.1-0.34) [0.23; 0.09]	

Table 8-6: ^{18}F-Fluoride PET; Whole-Body Analysis; Second Scan All Parameters; all patients; Descriptive statistics

8.1.3 Percentage Change (%Δ) Between Scans

12 patients had data for %Δ analysis (BCa-5, PCa-7). 11 patients had data for response analysis (BCa-5, PCa-6). 4 patients had a treatment response (including stable disease) maintained at 24-weeks (BCa-3(60%),PCa-1(16%))

8.1.3.1 %Δ SUV Parameters

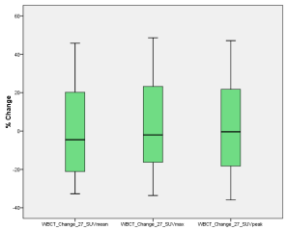
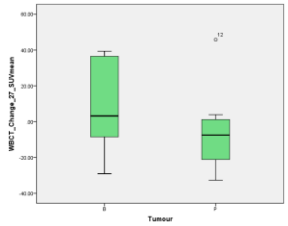
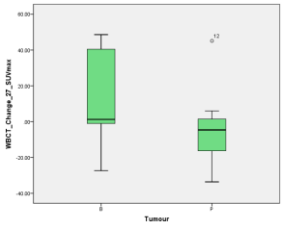
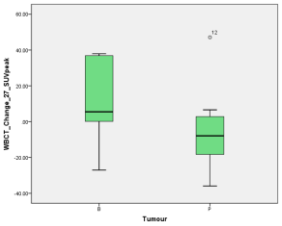
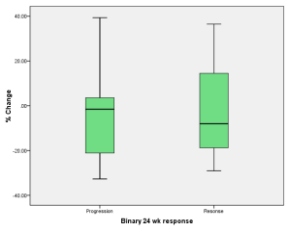
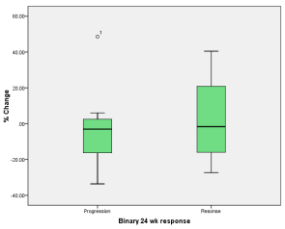
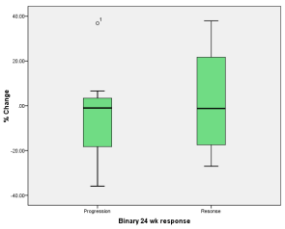
%ΔSUV$_{meanWB}$		%ΔSUV$_{maxWB}$		%ΔSUV$_{peakWB}$	
		Mean %Δ (Min-Max) [Median; SD]			
0.6 (-32.69-45.78) [-4.52; 26.87]		**3.28** (-33.61-48.6) [-2.01; 27.76]		**2.15** (-35.91-47.11) [-0.41; 26.69]	
					
Breast n=5	**Prostate** n=7	**Breast** n=5	**Prostate** n=7	**Breast** **n=5**	**Prostate** n=7
8.29 (-29.07-39.34) [3.2; 29.42]	**-4.89** (-32.69-45.78) [-7.5; 25.74]	**12.4** (-27.35-48.6) [1.27; 31.55]	**-3.24** (-33.61-45.11) [-4.62; 25.1]	**10.69** (-27.01-37.92) [5.5; 27.32]	**-3.94** (-35.91-47.11) [-7.91; 26.54]
					
p=0.639		p=0.343		p=0.432	
Progression n=7 (BCa 2; PCa 5)	**Response** n=4 (BCa 3; PCa 1)	**Progression** **n=7** **(BCa 2; PCa 5)**	**Response** n=4 (BCa 3; PCa 1)	**Progression** **n=7** **(BCa 2; PCa 5)**	**Response** n=4 (BCa 3; PCa 1)
-4.28 (-32.69-39.34) [-1.53; 23.87]	**-2.15** (-29.07-36.48) [-8.01; 27.6]	**-2.23** (-33.61-48.6) [-3.02; 26.22]	**2.45** (-27.35-40.47) [-1.67; 28.2]	**-4.25** (-35.91-36.86) [-0.99; 23.4]	**2.12** (-27.01-37.92) [-1.21; 27.34]
					
p=0.927		p=1.00		p=0.788	

Table 8-7: ^{18}F-Fluoride PET; Whole-Body Analysis; %Δ SUV Parameters; all patients; Descriptive statistics and response analyses; Mann-Whitney U tests for comparisons between distributions (B=Breast, P=Prostate)

The %Δ SUV parameters show more agreement between them than the baseline parameters (Tbl.8-7); there is no demonstrable significant difference between the distributions (Friedman's variance test, p=0.558).

Each response group contains patients with increases and decreases in the SUV parameter. The following waterfall plots (Figs.8-9,8-10,8-11) show the %Δ in the parameters for every patient.

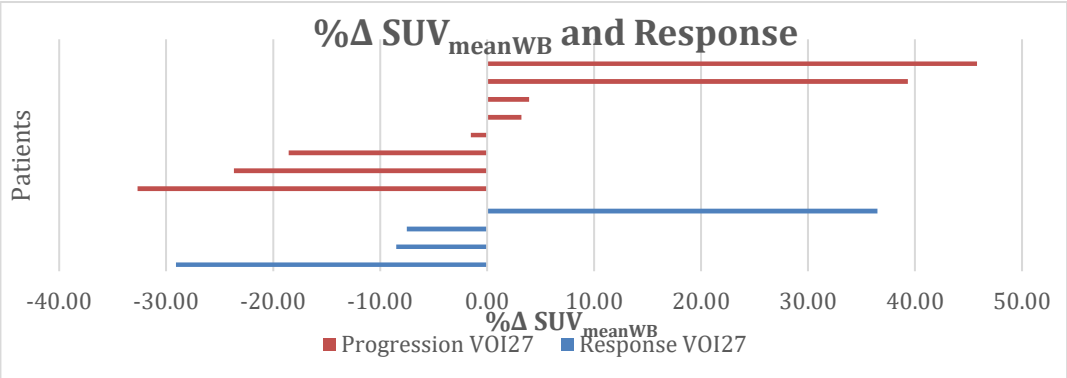


Figure 8-9: ^{18}F-Fluoride PET; Whole-Body Analysis; %ΔSUV$_{meanWB}^{VOI27}$; all patients split into response groups

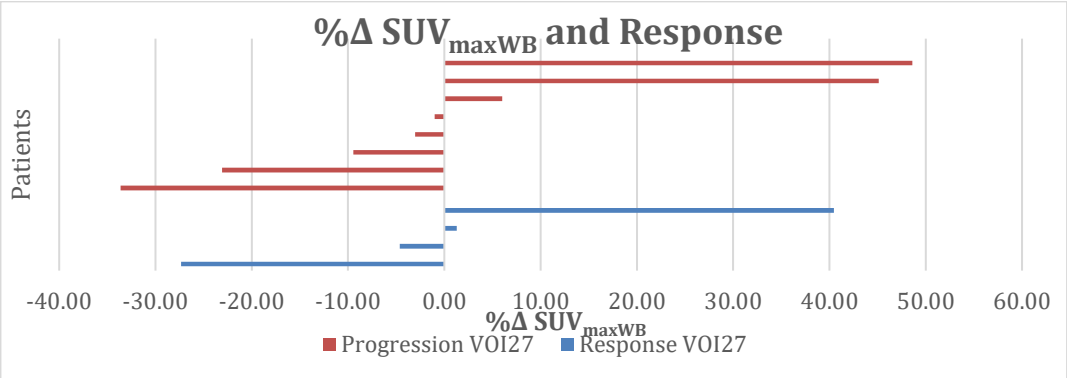


Figure 8-10: ^{18}F-Fluoride PET; Whole-Body Analysis; %ΔSUV$_{maxWB}^{VOI27}$; all patients split into response groups

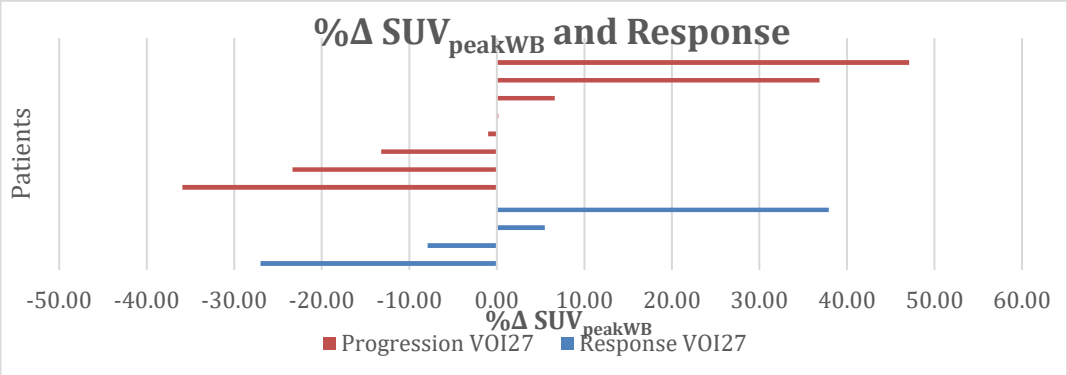


Figure 8-11: ^{18}F-Fluoride PET; Whole-Body Analysis; %ΔSUV$_{peakWB}^{VOI27}$; all patients split into response groups

There are no consistent findings from analysis of the %Δ SUV$_{WB}$ parameters. No utility has been shown for identifying a treatment response using ^{18}F-Fluoride PET 8-weeks after treatment initiation with these whole-body parameters (Mann-Whitney U tests to compare the response groups: %Δ SUV$_{mean/max/peak}$ p=0.927/1.0/0.788).

Response analysis of the breast cancer patients only can be reviewed in Table 8-8:

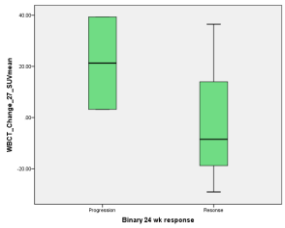
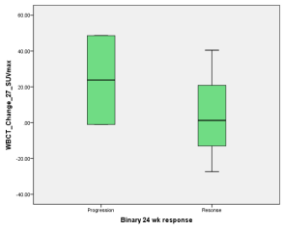
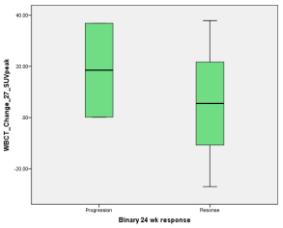
%ΔSUV$_{meanWB}$ Breast only		**%ΔSUV$_{maxWB}$ Breast only**		**%ΔSUV$_{peakWB}$ Breast only**	
Mean %Δ (Min-Max) [Median; SD]					
Progression n=2	**Response** n=3	**Progression** n=2	**Response** n=3	**Progression** n=2	**Response** n=3
21.27 (3.2-39.34) [21.27; 25.55]	**-0.37** (-29.07-36.48) [-8.51; 33.52]	**23.8** (-1-48.6) [23.8; 35.07]	**4.8** (-27.35-40.47) [1.27; 34.05]	**18.52** (0.18-36.86) [18.52; 25.94]	**5.47** (-27.01-37.92) [5.5; 32.47]
 p=0.4		 p=0.8		 p=1.0	

Table 8-8: ^{18}F-Fluoride PET; Whole-Body Analysis; %Δ SUV Parameters; Breast Patients Only; Descriptive statistics and response analyses; Mann-Whitney U tests for comparisons between distributions

None of the differences between the %Δ SUV parameters distributions reach statistical significance (%Δ SUV$_{mean/max/peak}$ p=0.4/0.8/1.0), but there is a suggestion of a greater increase in SUV between scans for those patients with PD by 24-weeks. A larger sample would be necessary to interrogate this further.

8.1.3.2 %Δ Volumetric Parameters

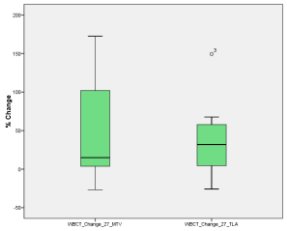
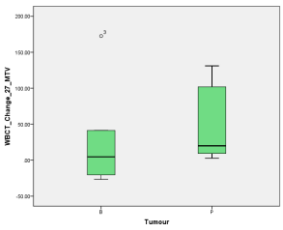
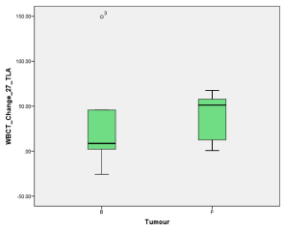
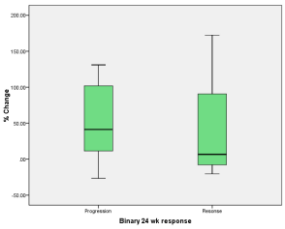
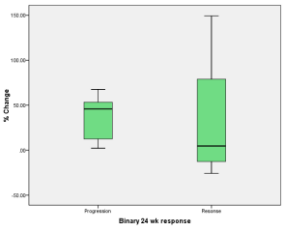
%ΔMTV$_{WB}$		%ΔTLA$_{WB}$	
Mean (Min-Max) [Median; SD]			
45.64 (-26.68-172.49) [14.92; 64.79]		**36.68** (-25.8-149.3) [31.98; 45.98]	
			
Breast n=5	**Prostate** n=7	**Breast** n=5	**Prostate** n=7
34.26 (-26.68-172.49) [4.61; 81.76]	**53.76** (2.8-131) [19.99; 55.27]	**36.02** (-25.8-149.3) [8.63; 68.28]	**37.16** (0.65-67.6) [51.23; 27.69]
			
p=0.432		p=0.530	
Progression n=7 (BCa 2; PCa 5)	**Response** n=4 (BCa 3; PCa 1)	**Progression** n=7 (BCa 2; PCa 5)	**Response** n=4 (BCa 3; PCa 1)
53.18 (-26.68-131) [41.28; 59.16]	**41.38** (-20.41-172.49) [6.71; 88.35]	**35.32** (2.16-67.6) [45.8; 25.88]	**33.2** (-25.8-149.3) [4.64; 78.79]
			
p=0.788		p=0.412	

Table 8-9: ^{18}F-Fluoride PET; Whole-Body Analysis; %Δ Volumetric Parameters; all patients; Descriptive statistics and response analyses; Mann-Whitney U tests for comparisons between distributions (B=Breast, P=Prostate)

There is no statistically significant difference in the distributions of the %Δ volumetric parameters between the tumour groups (Tbl.8-9) (Mann-Whitney U tests for comparison: %Δ MTV p=0.432, %ΔTLA p=0.530). More of the prostate patients had PD, but there is no significant difference in %ΔMTV or %ΔTLA between the response groups (p=0.788, 0.412 respectively).

There is heterogeneity of changes in the whole-body parameters within each response groups. The following charts plot the patient data for each response group (Figs.8-12,8-13). Most patients, irrespective of the response group, had an increase in the parameters between scans.

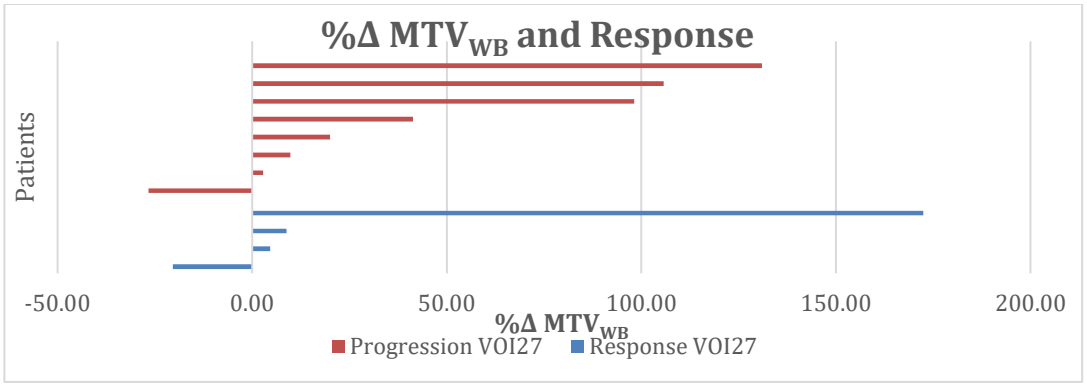


Figure 8-12: ^{18}F-Fluoride PET; Whole-Body Analysis; %ΔMTV$_{WB}^{VOI27}$; all patients split into response groups (Progression=Grey; Response=Blue)

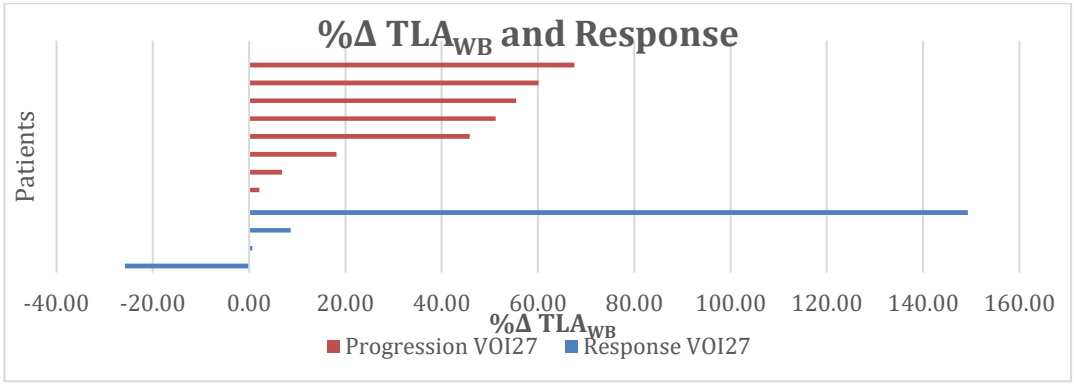


Figure 8-13: ^{18}F-Fluoride PET; Whole-Body Analysis; %ΔTLA$_{WB}^{VOI27}$; all patients split into response groups (Progression=Grey; Response=Blue)

The %Δ volumetric parameters have not shown utility for prediction treatment response at the per-lesion analysis (see earlier results chapter), nor per-patient target lesions analysis (see earlier results chapter). It is not clear why the volumetric parameters have increased for most patients, irrespective of the response group. The largest increase was in a patient deemed to have had a treatment response at 24-weeks; the increase in volume of ^{18}F-Fluoride uptake, as defined by the 27% iso-contour, may represents a flare response to treatment.

8.1.3.3 %Δ Heterogeneity Parameters

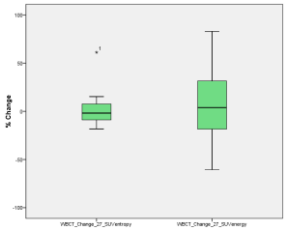
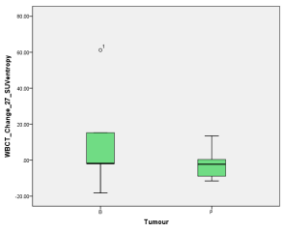
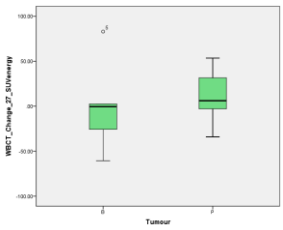
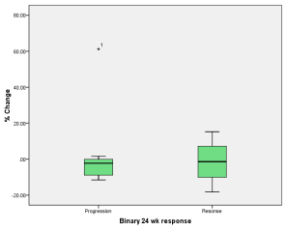
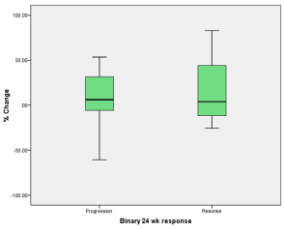
%ΔSUV$_{entropyWB}$		**%ΔSUV$_{energyWB}$**	
Mean (Min-Max) [Median; SD]			
3.1 (-18.24-61.21) [-1.86; 20.65]		**6.79** (-60.58-82.96) [3.9; 39.41]	
			
Breast n=5	**Prostate** n=7	**Breast** n=5	**Prostate** n=7
10.9 (-18.24-61.21) [-1.84; 30.51]	**-2.47** (-11.52-13.51) [-2.25; 8.62]	**-0.24** (-60.58-82.96) [-0.44; 52.96]	**11.81** (-34.11-53.47) [6.13; 30.13]
			
p=0.539		p=0.432	
Progression n=7 (BCa 2; PCa 5)	**Response** n=4 (BCa 3; PCa 1)	**Progression** n=7 (BCa 2; PCa 5)	**Response** n=4 (BCa 3; PCa 1)
4.21 (-11.52-61.21) [-2.25; 25.59]	**-1.44** (-18.24-15.26) [-1.39; 13.68]	**7.21** (-60.58-53.47) [6.13; 37.72]	**16.27** (-25.7-82.96) [3.9; 46.61]
			
p=1.0		p=1.0	

Table 8-10: ^{18}F-Fluoride PET; Whole-Body Analysis; %Δ SUV Heterogeneity Parameters; all patients; Descriptive statistics and response analyses; Mann-Whitney U tests for comparisons between distributions (B=Breast, P=Prostate)

There is no demonstrable statistically significant difference in the %Δ heterogeneity parameters between the tumour types (%ΔSUV$_{entropy}$ p=0.539, %ΔSUV$_{energy}$ p=0.432; Tbl.8-10), or response groups (p=1.0 for both parameters). The waterfall plots (Figs.8-14,8-15) show patients with increases and others with decreases in the parameters within each response group.

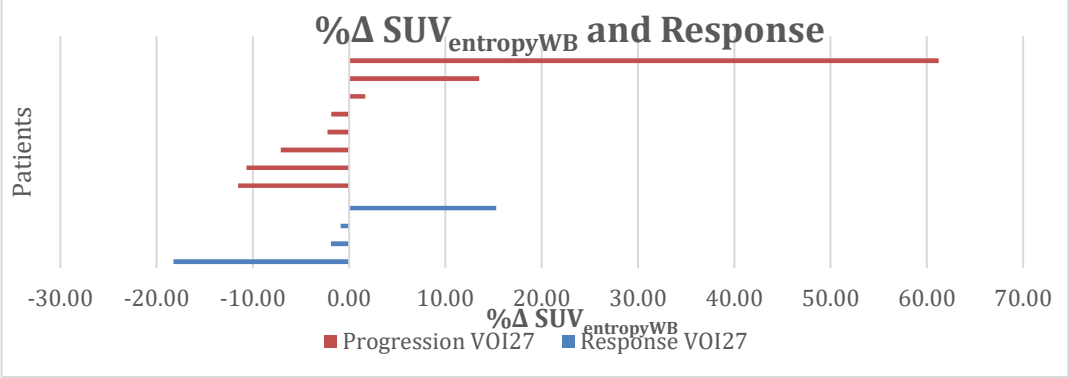


Figure 8-14: ^{18}F-Fluoride PET; Whole-Body Analysis; %ΔSUV$_{entropyWB}^{VOI27}$; all patients split into response groups

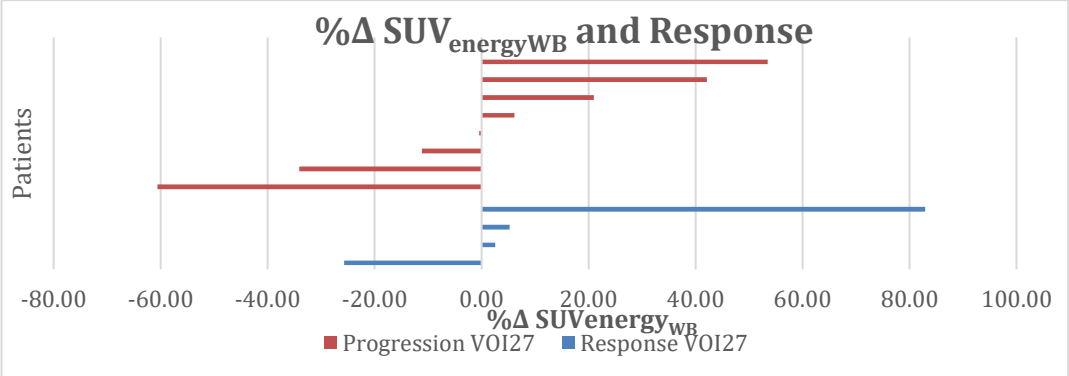


Figure 8-15: ^{18}F-Fluoride PET; Whole-Body Analysis; %ΔSUV$_{energyWB}^{VOI27}$; all patients split into response groups

Response analysis of the breast cancer patients only (Tbl.8-11) does not identify a statistically significant difference in the %Δ parameters between the response groups, although the patients with the greater increase in SUV$_{entropy}$ and greatest decrease in SUV$_{energy}$ between scan who did not have a sustained treatment response.

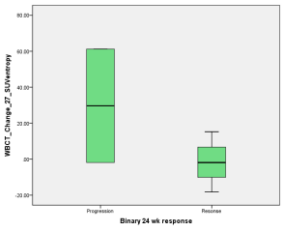
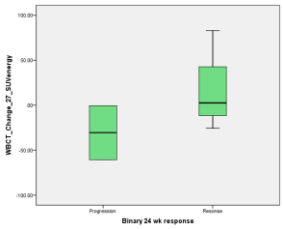
%ΔSUV$_{entropyWB}$ **Breast only**		**%ΔSUV$_{energyWB}$** **Breast only**	
Mean %Δ (Min-Max) [Median; SD]			
Progression **n=2**	**Response** n=3	**Progression** n=2	**Response** n=3
29.68 (-1.84-61.21) [29.68; 44.58]	-1.62 (-18.24-15.26) [-1.88; 16.75]	-30.51 (-60.58--0.44) [-30.51; 42.52]	19.93 (-25.7-82.96) [2.54; 56.38]
			
p=0.4		p=0.4	

Table 8-11: ^{18}F-Fluoride PET; Whole-Body Analysis; %Δ SUV Heterogeneity Parameters; Breast Patients Only; Descriptive statistics and response analyses; Mann-Whitney U tests for comparisons between distributions

8.1.3.4 %Δ Uptake Rate Constant (Ki_{WB})

Only 8 patients had data suitable for %ΔKi analysis, 4 breast cancer patients (2 with a treatment response) and 4 prostate cancer patients (1 with a treatment response).

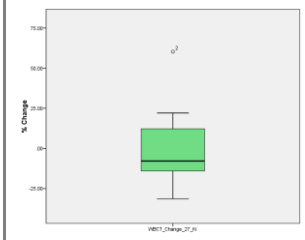
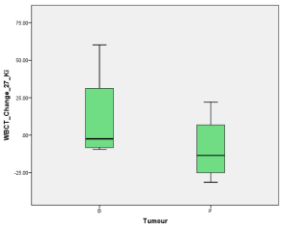
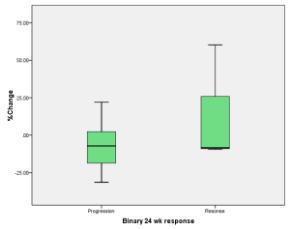
%ΔKi$_{WB}$ **Mean** (Min-Max) [Median; SD]				
All pts	**Breast** n=4	**Prostate** n=4	**Progression** n=5 (BCa 2; PCa 3)	**Response** n=3 (BCa 2; PCa 1)
1.19 (-31.42-60.24) [-7.82 ;28.42]	**11.47** (-9.47-60.24) [-2.44; 32.91]	**-9.09** (-31.42-22.08) [-13.52; 22.81]	**-6.56** (-31.42-22.08) [-7.18; 20.38]	**14.1** (-9.47-60.24) [-8.46; 39.96]
	 p=0.343		 p=0.786	

Table 8-12: ^{18}F-Fluoride PET; Whole-Body Analysis; %Δ Ki; all patients; Descriptive statistics and response analyses; Mann-Whitney U tests for comparisons between distributions (B=Breast, P=Prostate)

The Ki^{WB} distribution seems higher in those with a treatment response(Tbl.8-12), but this is not confirmed statistically (p=0.343). The waterfall plot (Fig.8-16) demonstrates the individual patient data; the change in Ki^{WB} is inconsistent and does not appear to predict treatment response in these patients. There are too few patients for response analysis of the breast cancer patients only.

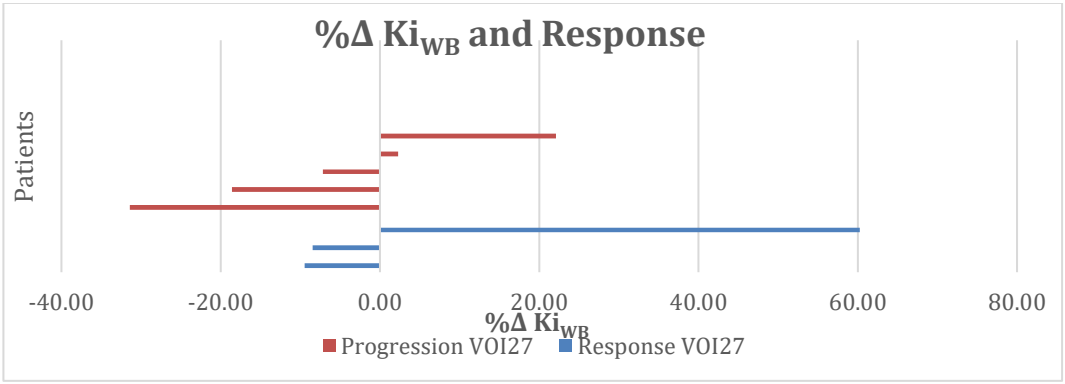


Figure 8-16: ^{18}F-Fluoride PET; Whole-Body Analysis; %ΔKi_{WB}^{VOI27}; all patients split into response groups

8.2 OS Analysis

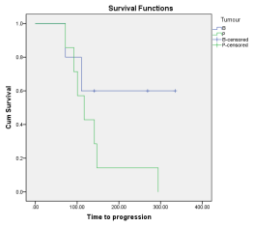
As previously discussed, there are too few deaths in this data to warrant meaningful OS analysis.

8.3 PFS Analysis

PFS analysis with Cox regression analysis is more valid than OS analysis in these patients because of the increased number of events.

12 patients had PFS data available for analysis, 5 breast cancer patients and 7 prostate cancer patients. 2 breast cancer patients had PD at the point of data analysis, and 7 prostate cancer patients.

8.3.1 Tumour Group Analysis



Breast cancer (B) - blue line
Prostate cancer (P) - green line
p=0.153

Figure 8-17: ^{18}F-Fluoride PET; Whole-Body Analysis; KM PFS Analysis of tumour groups (Log rank analysis; B=Breast cancer; P=Prostate cancer)

There is no statistically significant PFS benefit for the breast cancer patients has been identified (Fig.8-24), although this most likely represents the small number of patients and events.

8.3.2 Baseline Scan PFS Analysis

8.3.2.1 SUV_{WB} Parameters

The results of the Cox analysis (Tbl.8-19) show is no demonstrable correlation between the baseline SUV_{WB} parameters and the risk of progression:

	SUV_{meanWB}	**SUV_{maxWB}**	**SUV_{peakWB}**
p-value [HR]	0.932 [1.002]	0.84 (1.003)	0.75 (1.006)

Table 8-13: ^{18}F-Fluoride PET; Whole-Body Analysis; Baseline Scan SUV Parameters; Cox regression PFS Analysis; All Patients

Table 8-20 reports the KM PFS analyses. Using an SUV_{meanWB}/SUV_{maxWB}/SUV_{peakWB} cut-off at 12/21.08/16.6 (respectively) the same 3 patients have been selected, creating identical KM analysis for all three parameters. There is significant separation of the PFS curves (p=0.037), with a 55-day median-PFS benefit for the patients with the higher baseline SUV parameters. Analysis of the prostate cancer patients only does not demonstrate a predictive association with these baseline SUV parameters (SUV_{meanWB}/SUV_{maxWB}/SUV_{peakWB} p=0.475, 0.909, 0.153 respectively). The analysis of all the patients together is most likely influenced by the tumour subgroup, and an association between these parameters and PFS cannot be concluded.

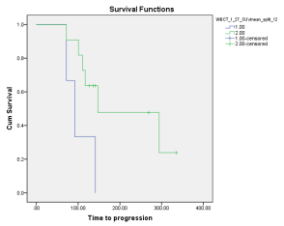
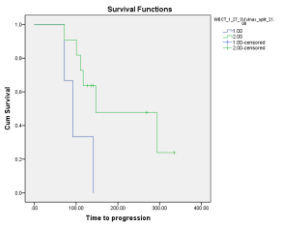
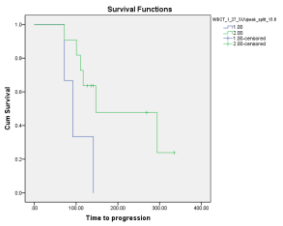
	SUV_{meanWB}	SUV_{maxWB}	SUV_{peakWB}
	Median PFS in days (95% CI)		
			
Cut-off	12.0	21.08	15.6
≤ (1)	92 (59.99-124.01)	92 (59.99-124.01)	92 (59.99-124.01)
> (2)	147 (0-300.12)	147 (0-300.12)	147 (0-300.12)
p-Value (Log Rank)	0.037	0.037	0.037

Table 8-14: ^{18}F-Fluoride PET; Whole-Body Analysis; Baseline Scan SUV Parameters; KM PFS Analysis; All Patients

8.3.2.2 Whole-Body SUV Volumetric Parameters

Cox regression analysis (Tbl.8-21) does not demonstrate a correlation between the size of the baseline whole-body volumetric parameters and PFS in these patients.

	MTV_{WB}	TLA_{WB}
p-value [HR]	0.683 (1)	0.794 (1)

Table 8-15: ^{18}F-Fluoride PET; Whole-Body Analysis; Baseline Scan Volumetric Parameters; Cox regression PFS Analysis; All Patients

The data has been dichotomised for KM analysis – see Tbl.8-22:

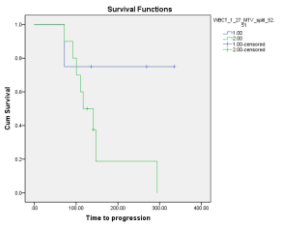
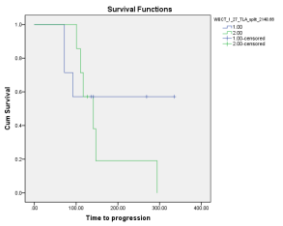
	MTV_{WB}	TLA_{WB}
	Median PFS in days (95% CI)	
		
Cut-off	52.51	2148.66
≤ (1)	Not reached	Not reached
> (2)	117 (74.3-159.7)	141 (91.77-190.23)
p-Value (Log Rank)	0.137	0.445

Table 8-16: ^{18}F-Fluoride PET; Whole-Body Analysis; Baseline Scan Volumetric Parameters; KM PFS Analysis; All Patients

The KM analysis does not identify significant separation of the PFS curves. Of the 4 patients with an MTV_{WB} less than 52.5cm^3 only one had progressed at the time of data analysis, suggesting a lower tumour burden at baseline is associated with a prolonged PFS; however, the log rank analysis of the curves is not significant, and the numbers are small.

Analysis of the prostate cancer patients (Tbl.8-23) shows a significant separation of the PFS curves for MTV_{WB} (p=0.048) shows a PFS advantage for patients with larger MTV_{WB} measurements. No significant separation is identified with TLA_{WB}. The per-lesion analysis did not identify an association between baseline MTV and PFS. This might suggest the whole-body

quantification has unveiled prognostic information about these patients, not shown from individual lesion analysis; perhaps a larger burden of osteoblastic mediated prostate cancer bone metastases indicates a different disease biology with a slower progression trajectory. With this small sample, this might simply represent a statistical anomaly.

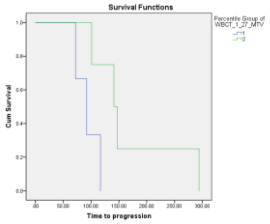
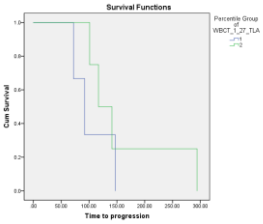
	MTV_{WB} **Prostate only**	**TLA_{WB}** **Prostate only**
	Median PFS in days (95% CI)	
		
Dichotomisation	108.56	4212.65
≤ (1)	92 (60-124)	92 (60-124)
> (2)	141 (95.9-186.1)	117 (77.8-156.2)
p-Value (Log Rank)	0.048	0.400

Table 8-17: ^{18}F-Fluoride PET; Whole-Body Analysis; Baseline Scan Volumetric Parameters; KM PFS Analysis; Prostate patients only

8.3.2.3 SUV_{WB} Heterogeneity Parameters

The Cox regression analysis below does not identify a significant correlation between the magnitude of the baseline SUV_{WB} heterogeneity parameters and the PFS of these patients (Tbl.8-24).

	$SUV_{entropyWB}$	**$SUV_{energyWB}$**
p-value [HR]	0.728 [0.794]	0.439 [79.315]

Table 8-18: ^{18}F-Fluoride PET; Whole-Body Analysis; Baseline Scan SUV Heterogeneity Parameters; KM PFS Analysis; All Patients

The data has been dichotomised for KM analysis (Tbl.8-25).

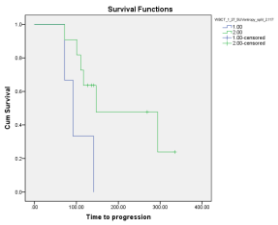
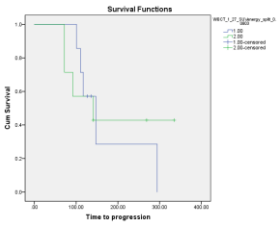
	$SUV_{entropyWB}$	**$SUV_{energyWB}$**
	Median PFS in days (95% CI)	
		
Dichotomisation	2.12	0.0803
≤ (1)	92 (59.99-124.01)	147 (101.18-192.82)
> (2)	147 (0-300.12)	141 (15.25-266.75)
p-Value (Log Rank)	0.037	0.842

Table 8-19: Median PFS times from KM analysis with optimised dichotomisation; Whole-Body Heterogeneity Parameters; ^{18}F-Fluoride PET; Baseline Scan; WBCT Analysis; All Patients; VOI_{27} only.

There is a significant separation of the PFS curves using $SUV_{entropyWB}^{VOI27}$ – patients with $SUV_{entropyWB}^{VOI27}$ greater than 2.12 have a median PFS benefit of 55 days (p=0.037). Patients with more a more heterogeneous whole-body measurement have a better PFS. This is not

maintained after analysis of only the prostate cancer patients (p=0.909); $SUV_{entropy}$ has not been shown to be an independent predictor of treatment response.

There is no significant separation of the survival curves for $SUV_{energyWB}^{VOI27}$ (p=0.842 for all patients, p-0.909 for prostate cancer patients only, KM analysis).

8.3.2.4 Ki_{WB}

n=11, 8 events

	Ki_{WB}
p-value [HR]	0.747 [3.168]

Table 8-20: ^{18}F-Fluoride PET; Whole-Body Analysis; Baseline Scan Ki; Cox regression PFS Analysis; All Patients

Cox regression univariate analysis demonstrates no correlation between the Ki_{WB} and the risk of PD (Tbl.8-26). The data has been dichotomised for KM analysis (Tbl.8-27).

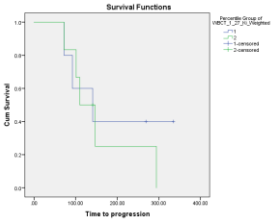
	Ki_{WB}
	Median PFS in days (95% CI
KM Plot	
Median	0.25
≤ Median (1)	141.00 (35.79-246.21)
> Median (2)	110.00 (65.83-154.17)
p-Value (Log Rank)	0.636

Table 8-21: ^{18}F-Fluoride PET; Whole-Body Analysis; Baseline Scan Ki; KM PFS Analysis; All Patients

No significant separation of the PFS curves (p=0.636 for all patients, p=0.832 for prostate cancer patients only). Ki_{WB} has shown no PFS prognostic utility.

8.3.3 %Δ Whole-Body Parameters - PFS Analysis

11 patients had data suitable for analysis with the %Δ parameters, 5 breast and 6 prostate cancer patients. 8 patients had PD at the time of data analysis (2 breast cancer patients and all 6 prostate cancer patients.

8.3.3.1 %ΔSUV$_{WB}$ Parameters

Table 8-28 details the outcomes from Cox regression analyses. There is no demonstrable correlation between the PFS and the size of change in the SUV$_{WB}$ parameters.

	%Δ SUV$_{meanWB}$	%Δ SUV$_{maxWB}$	%Δ SUV$_{peakWB}$
p-value [HR]	0.388 (0.985)	0.398 (0.986)	0.33 (0.984)

Table 8-22: ^{18}F-Fluoride PET; Whole-Body Analysis; %Δ SUV Parameters; Cox regression PFS Analysis; All Patients

The data has been dichotomised for KM analysis. There is no statistically significant separation of the PFS curves (Tbl.8-29).

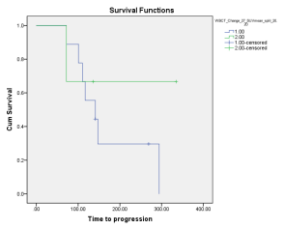
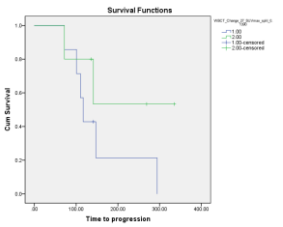
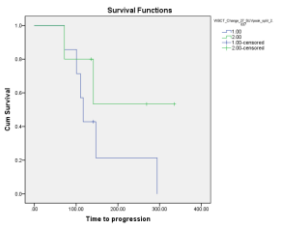
	%ΔSUV$_{meanWB}$	%ΔSUV$_{maxWB}$	%ΔSUV$_{peakWB}$
	Median PFS in days (95% CI)		
			
Cut-off	20.20	0.139	2.837
≤ (1)	141 (70.88-211.12)	117 (99.04-134.96)	117 (99.04-134.96)
> (2)	Not reached	Not reached	Not reached
p-Value (Log Rank)	0.364	0.227	0.227

Table 8-23: ^{18}F-Fluoride PET; Whole-Body Analysis; %Δ SUV Parameters; KM PFS Analysis; All Patients

No statistically significant separation of the PFS curves has been identified using the %Δ SUV parameters. The patients with higher positive changes in the SUV parameters were those with longer PFS.

Table 8-30 documents KM analysis of the prostate patients only, showing a statistically significant separation of the PFS curves only with %ΔSUV$_{meanWB}$ (p=0.025), with a 46-day median PFS benefit for patients with an increase in SUV$_{meanWB}$ between scans. This is not replicated with the other %Δ SUV parameters. The per-lesion analysis did not demonstrate this association for the prostate cancer lesions. An increase in SUV mean might suggest a flare response in patients, which could suggest a flare response. There were too few prostate cancer patients to compare the %ΔSUV$_{meanWB}$ between the response groups, but an increase in SUV$_{meanWB}$ was suggested in the breast cancer patients with a treatment response, and there was a statistical trend of this association with per-lesion analysis.

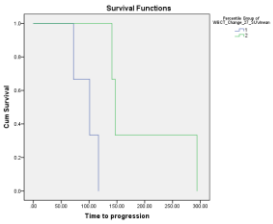
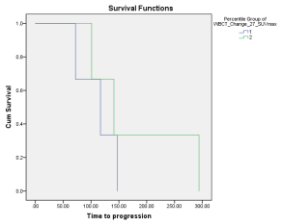
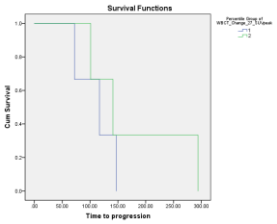
	%ΔSUV$_{meanWB}$ Prostate only	**%ΔSUV$_{maxWB}$ Prostate only**	**%ΔSUV$_{peakWB}$ Prostate only**
Median PFS in days (95% CI)			
KM Plot			
Cut-off	-7.50%	-4.62%	-7.91%
≤ (1)	101 (54.6-147.4)	117 (45-189)	117 (45-189)
> (2)	147 (137.4-156.6)	141 (77-205)	141 (77-205)
p-Value	0.025	0.486	0.486

Table 8-24: ^{18}F-Fluoride PET; Whole-Body Analysis; %Δ SUV Parameters; KM PFS Analysis; Prostate patients only

8.3.3.2 %Δ WB Volumetric Parameters

Table 8-31 details the outcomes from Cox regression analyses. There is no demonstrable statistically significant association between the size of change of the volumetric parameters and PFS.

	%ΔMTV$_{WB}$	**%ΔTLA$_{WB}$**
p-value [HR]	0.607 (1.003)	0.947 (0.999)

Table 8-25: ^{18}F-Fluoride PET; Whole-Body Analysis; %Δ Volumetric Parameters; Cox regression PFS Analysis; All Patients

The data has been dichotomised for KM analysis (Tbl.8-32). There is no statistically significant separation of the PFS curves (%ΔMTV$_{WB}$ p=0.151, %ΔTLA$_{WB}$ p=0.304). There is the suggestion of a PFS advantage for those patients with a smaller increase (<41%) and decrease in MTV between scans, and perhaps with a larger study group this might have been confirmed.

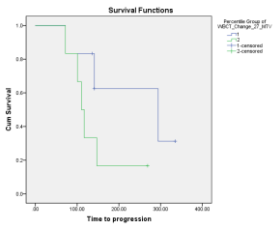
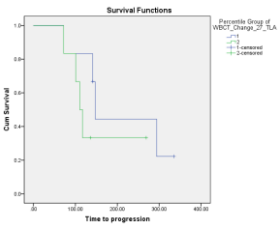
	%ΔMTV$_{WB}^{VOI27}$	**%ΔTLA$_{WB}^{VOI27}$**
		
Cut-off	41.28%	45.8%
≤ (1) [Days (95% CI)]	294 (58.5-529.5)	147.00 (135.24-158.76)
> (2) [Days (95% CI)]	110 (90.8-129.2)	110.00 (90.80-129.20)
p-Value (Log Rank)	0.151	0.304

Table 8-26: ^{18}F-Fluoride PET; Whole-Body Analysis; %Δ Volumetric Parameters; KM PFS Analysis; All Patients

The analysis for the prostate cancer patients is in Table 8-33 showing a trend towards a PFS benefit for patients with a smaller increase (or decrease) in MTV$_{WB}$ between scans (p=0.198), and a statistically significant 46-day median PFS benefit for the prostate patients with a smaller

increase (or decrease) in TLA_{WB} between scans. This is despite an increase in SUV between scans being associated with an improved PFS, and most likely reflects the impact of MTV on the calculation of TLA. With this small sample group, however, this association would need to be confirmed with a larger sample.

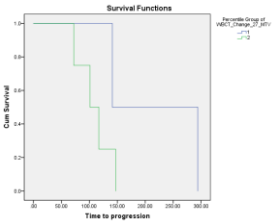
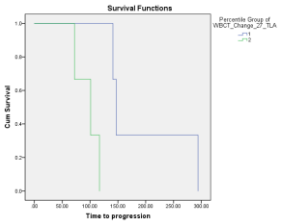
	$\%\Delta MTV_{WB}^{VOI27}$ Prostate only	**$\%\Delta TLA_{WB}^{VOI27}$ Prostate only**
		
Median	+19.99%	+51.23%
≤ (1) [Days (95% CI)]	141 (n/a)	147 (137.4-156.6)
> (2) [Days (95% CI)]	101 (56.9-145.1)	101 (54.6-147.4)
p-Value (Log Rank)	0.198	0.025

Table 8-27: ^{18}F-Fluoride PET; Whole-Body Analysis; %Δ Volumetric Parameters; KM PFS Analysis; All Patients

8.3.3.3 $\%\Delta SUV_{WB}$ Heterogeneity Parameters

The Cox regression analysis of this data does not a significant correlation between PFS and the size of change of the heterogeneity parameters (Tbl.8-34).

	$\%\Delta SUV_{entropyWB}$	**$\%\Delta SUV_{energyWB}$**
p-value [HR]	0.566 [1.014]	0.829 [0.998]

Table 8-28: ^{18}F-Fluoride PET; Whole-Body Analysis; %Δ SUV Heterogeneity Parameters; Cox regression PFS Analysis; All Patients

The data for has been dichotomised for KM analysis (Tbl.8-35).

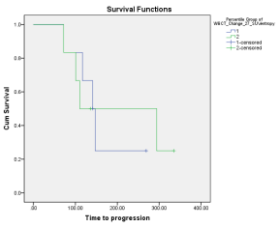
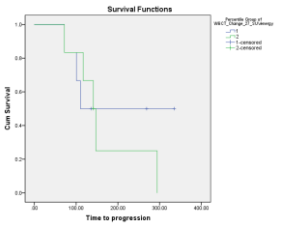
	$\%\Delta SUV_{entropyWB}^{VOI27}$	**$\%\Delta SUV_{energyWB}^{VOI27}$**
KM Plots		
Dichotomisation	-1.86%	+3.9%
Low ≤ (1)	141 (112.2-169.8)	110 (n/a)
High > (2)	110 (0-295.3)	141 (112.2-169.8)
p-Value (Log Rank)	0.906	0.637

Table 8-29: ^{18}F-Fluoride PET; Whole-Body Analysis; %Δ SUV Heterogeneity Parameters; KM PFS Analysis; All Patients

No significant separation of the PFS curves is identified with either parameter ($\%\Delta SUV_{entropyWB}$ p=0.906, $\%\Delta SUV_{energyWB}$ p=0.637), and analysis of the prostate cancer patients only still identifies no PFS prognostic utility (p=0.486 and 0.441 respectively).

8.3.3.4 %ΔKi$_{WB}$

8 patients had data suitable for analysis (4 prostate, 4 breast). 6 patients had PD, all 4 prostate patients and 2 of the breast patients.

The Cox regression analysis (Tbl.8-36) does not identify any correlation between the magnitude of %ΔKi$_{WB}$ and PFS in these patients.

	%ΔKi$_{WB}$
p-value [HR]	0.183 [0.967]

Table 8-30: Cox regression PFS Analysis; %Δ Whole-Body Ki; ^{18}F-Fluoride PET; Percentage Change; WBCT Analysis; All Patients; all threshold methods.

The data has been dichotomised for KM analysis (Tbl.8-37), but no statistically significant separation of the survival curves has been identified in these patients.

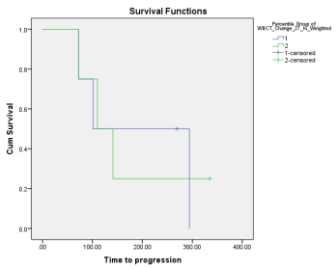
	%ΔKi$_{WB}$
KM Plot	
Median	-7.82%
≤ Median (1) **[Days (95% CI)]**	101.00 (0-246.04)
> Median (2) **[Days (95% CI)]**	110.00 (42.38-117.62)
p-Value (Log Rank)	0.933

Table 8-31: ^{18}F-Fluoride PET; Whole-Body Analysis; %Δ Ki; KM PFS Analysis; All Patients

Analysis of the prostate cancer patients (n=4) can be seen below (Tbl.8-38). The prostate patients with an increase in Ki$_{WB}$ between scans show a trend towards a significant PFS benefit (p=0.09), but with so few patients conclusions cannot be definitively drawn.

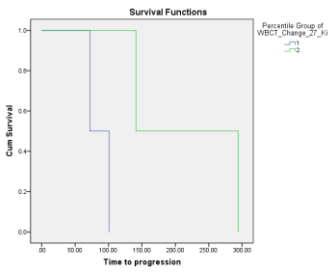
	%ΔKi$_{WB}$ Prostate Only
	
Median	-13.51%
p-Value (Log Rank)	0.09

Table 8-32: ^{18}F-Fluoride PET; Whole-Body Analysis; %Δ Ki; KM PFS Analysis; Prostate patients only

8.4 Discussion

Baseline Parameter	SUV_{meanWB}	SUV_{maxWB}	SUV_{peakWB}	MTV_{WB}	TLA_{WB}	$SUV_{entropyWB}$	$SUV_{energyWB}$	Ki_{WB}
Tumour	P>B p=0.298	P>B =0.147	P>B p=0.147	P>B p=0.298	P>B p=0.042	P>B p=0.190	P<B p=0.083	X
Response	X	X	X	X	X	X	X	X
↑PFS	X	X	X	All pts X ↑ P only p=0.048	X	X	X	X

%Δ Parameter	SUV_{meanWB}	SUV_{maxWB}	SUV_{peakWB}	MTV_{WB}	TLA_{WB}	$SUV_{entropyWB}$	$SUV_{energyWB}$	Ki_{WB}
Tumour	X	X	X	X	X	X	X	X
Response	X	X	X	X	X	X	X	X
↑PFS	↑ P only p=0.025	X	X	↓ P only p=0.198	↓ P only p=0.025	X	X	↑ P only p=0.09

Table 8-33: ^{18}F-Fluoride PET; Whole-Body Analysis; Summary Tables

This study was designed to explore the potential predictive and prognostic utility from quantifying the whole-skeletal burden of ^{18}F-Fluoride avid bone metastases. Per-lesion analysis showed clinically significant information can be gleaned from the baseline parameters, and from parameter changes following treatment, but inter-lesion heterogeneity was evident. The most clinically relevant lesion(s) for analysis is unclear; whole-body quantification perhaps results in parameters taking this heterogeneity into account, and this might have predictive and prognostic relevance. Whole-body qualitative assessment of bone scans and PET imaging is routine in clinical practice; providing quantification parameters might standardise analysis, and demonstrate changes not identified visually.

There is limited literature reporting similar approaches to whole-body ^{18}F-Fluoride PET quantification. When this study was conceived, I was only aware of one study exploring whole-body quantification. Abouzied et al. reported in a meeting abstract a study of 25 patients with varied malignancies, but confirmed bone metastases. A Hounsfield unit threshold of 150 was used to segment the skeleton from the CT, and this was used as a mask to segment bone metastases from the ^{18}F-FDG PET, segmenting with a 50% iso-contour [188]. They demonstrated feasibility of this approach, but they did not report clinical outcome analysis.

More recently further data has been published addressing whole-body PET quantification, but primarily focusing on volumetric parameters. Building on the knowledge of skeletal burden of metastatic disease predicting survival of prostate cancer patients [210], ^{18}F-Fluoride PET imaging has been interrogated for the same purpose.

Etchebehere reported a study of 42 patients treated with Alpharadin; baseline tumour burden identified on ^{18}F-Fluoride PET imaging was highly and significantly correlated with OS (HR 5.990, p=0.02)[204]. Visual analysis, and whole-body SUV_{mean} and SUV_{max} parameters were not predictive of OS or PFS, but the average SUV predicted for skeletal related events.

Most recently, Lindgren Belal et al. have shown correlation between whole-skeletal metastatic burden metrics with comparable bone scan indices, and significant correlation with OS [205]. Wassberg et al., in a study of repeatability of ^{18}F-Fluoride PET for prostate cancer bone metastases, measured the whole-skeletal ^{18}F-Fluoride avid metastatic burden, using a 50% iso-contour, and calculated a parameter analogous to the TLA used in this thesis study. They repeated the PET imaging within 8 days, and demonstrated reasonable repeatability coefficients for the whole-body parameters (23% for TLA parameter, 35% for skeletal volumetric burden), and reasonable performance of the per-lesion metrics of SUV_{max} and SUV_{mean} (26% and 24%, respectively)[211].

To my knowledge this this is the first study examining the potential clinical utility of whole-body ^{18}F-Fluoride PET quantification for bone metastases using baseline and %Δ parameters.

The potential clinical utility of ^{18}F-Fluoride PET has been suggested in the per-lesion analyses of this thesis. The per-lesion analysis demonstrated the different ^{18}F-Fluoride avidity of breast and prostate cancer bone metastases; univariate analysis of the patients may be misleading, but the number of patients in this study has prevented meaningful tumour subgroup analyses. There is, however, evidence of meaningful and clinically relevant image signal.

Whole-skeletal quantification shows a trend for the prostate cancer metastases to have higher baseline SUV parameters than the breast patients, have a greater ^{18}F-Fluoride avid bone metastasis burden (MTV), and have whole-body image heterogeneity tending away from homogeneity (higher $SUV_{entropy}$ and lower SUV_{energy}). Prostate cancer metastases tend to display predominantly osteoblastic activity compared to the more mixed osteoblastic and osteoclastic qualities of breast cancer deposits. This would account for this higher ^{18}F-Fluoride avidity. It is not clear what feature of the bone metastases is accounting for the differences in the heterogeneity parameters.

The differences in SUV and volumetric parameters were also identified with the per-lesion analysis of these patients. The whole-skeletal quantification technique is providing patient-specific data discriminating features of the underlying bone pathology. Unfortunately, this has not provided the ability to predict a treatment response using the baseline scans; no utility was identified with per-lesion analysis. There may be no predictive information to be extracted from the baseline scans, but certainly the whole-skeletal quantification method has not yielded further discriminatory value.

The per-lesion analysis of the change in the parameters between the scans suggested an association between an increased SUV and a sustained treatment response for the breast cancer lesions. The whole-skeletal quantification has suggested this pattern, but this has not reached statistical significance (p=0.4 for SUV_{meanWB}). The statistical benefit of the per-lesion analysis

enables significance to be concluded for smaller differences; further interrogation of the predictive value of the %Δ SUV_{WB} parameters with a larger study group is indicated.

No other %Δ parameter has suggested utility in these patients for predicting the clinical treatment response.

An increase in SUV_{mean}^{WB} between scans is associated with a 46-day median PFS benefit in these prostate cancer patients (p=0.025) suggesting a flare response of increased osteoblastic activity in healing bone. This was not identified with the per-lesion or per-patient target-lesion analyses, and may suggest a benefit from quantification of the whole-skeleton metastatic burden. Similarly, a trend towards a PFS benefit from a decrease in the MTV of prostate cancer patients (p=0.198), and the significant benefit from a decrease in TLA (p=0.025) was not identified with per-patient target-lesion analysis. This may suggest prognostic benefit from the quantification of the whole-skeletal metastatic burden.

Segmentation of the skeletal PET data is readily achievable using freely available software. The 27% iso-contour was used to segment the skeletal uptake to select 'malignant' bone, but this may not be the most discriminatory method. Certainly, for a per-lesion analysis an iso-contour is individualised for each lesion, but for a whole-body segmentation this is likely to be less specific for identifying malignancy. A computational segmentation, such as FLAB, might be preferable, but such an approach for whole-body analysis is currently hindered by computational power. The whole-body quantification is likely to be impacted by random image noise; individual voxels of higher uptake are not necessarily seen on qualitative assessment, but may alter the parameters. In this thesis, I wanted to include all segmented voxels, and applied a volumetric weighting to limit the impact of these small lesions. An alternative method might be to exclude VOIs below a certain size. Smaller VOIs, even if truly representing malignant bone, are more prone to partial volume effects and this may degrade the quantification. The repeatability and reproducibility of these whole-body parameters has not been included in this thesis, but would be an essential step before expanding investigation of the clinical utility. ^{18}F-Fluoride may not be the most clinically predictive or prognostic PET tracer; the same skeletal segmentation methodology has been used to analyse the initial patients recruited to the FAB-P and FAB-B studies, enabling whole-body skeletal quantification of ^{11}C-Choline and ^{18}F-FDG PET. There were too few patients analysed for inclusion in this thesis, but the increased extra-skeletal physiological uptake of these tracers makes a methodology to segment just the skeletal PET data an attractive proposition.

One key area of data yet to be fully analysed is generation of skeletal burden metrics; the CT skeletal mask can be analysed similarly to the PET scan to generate 'whole-skeletal' volumes against which the ^{18}F-Fluoride avid volume could be compared. Other groups have suggested prognostic utility of this approach, and this might provide validation of this whole-body analysis methodology.

In summary, this exploratory study has demonstrated a readily applicable method for selecting only skeletal PET data for analysis, by applying a mask created from the CT component of the scan. The predictive and prognostic utility of ^{18}F-Fluoride PET scans suggested from per-lesion analysis has not been statistically confirmed using the whole-body parameters measured in this

study, but this has been impacted by the relatively small study population. There is a strong suggestion of discriminatory signal being quantified using these whole-body parameters, and this certainly warrants further investigation as the trials approach full recruitment.

Chapter 9 DW-MRI Per-lesion Analysis – Response assessment using DW-MRI quantification parameters of patients with bone metastases from breast or prostate cancer

The methods of this chapter and the statistical approaches for this section can be reviewed in Chapter 2.3.8.

9.1 Tumour Groups

92 lesions are included from 20 patients (BCa-5, PCa-15). 82 lesions (18 patients) had data for response analysis (BCa-22, PCa-60). 23 lesions were from patients with a treatment response (including stable disease) (BCa-12(55%), PCa-11(18%)). 59 lesions were from patients with PD (BCa-10(45%), PCa-49(82%)).

9.2 Baseline Scan

9.2.1 ADC Parameters

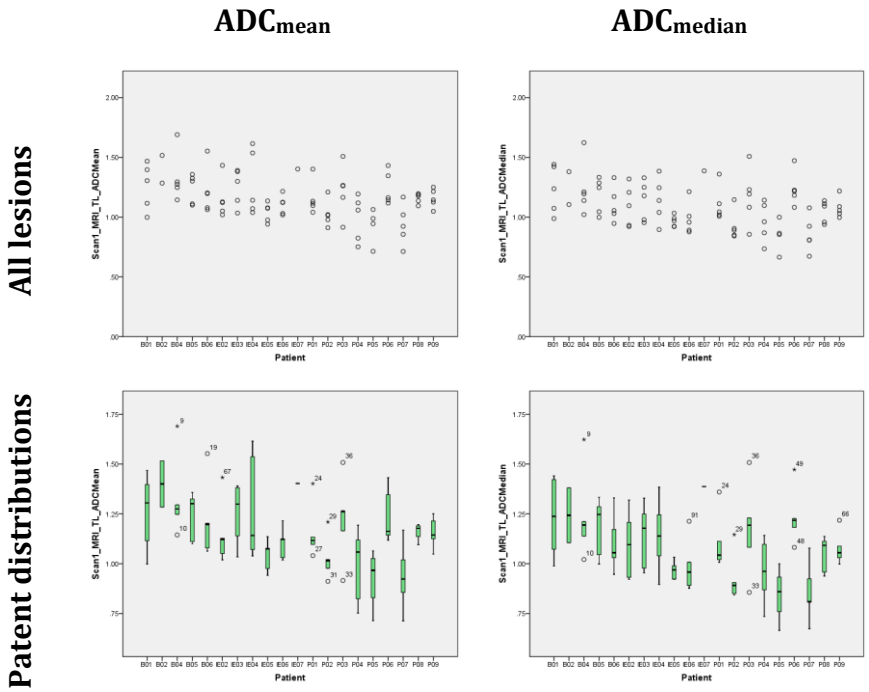


Figure 9-1: DW-MRI Per-lesion Analysis; Baseline ADC Parameters (ADC_{mean}-left, ADC_{median}-right); TOP – all lesions for all patients; BOTTOM – ADC distributions for all patients (ADC expressed as 10-3 mm^2/s) (B - Breast cancer patients; P – Prostate cancer patients)

Figure 9-1 shows the inter-lesion heterogeneity, both within and between patients. Table 9-1 summarises the parameter distributions for all lesions, including the sub-group analyses of tumour group and treatment response.

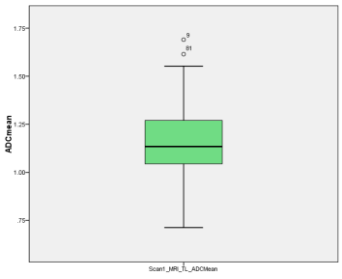
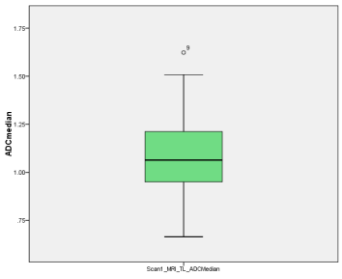
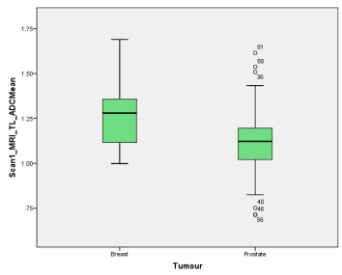
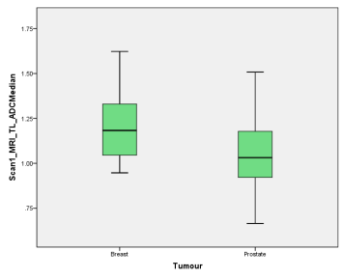
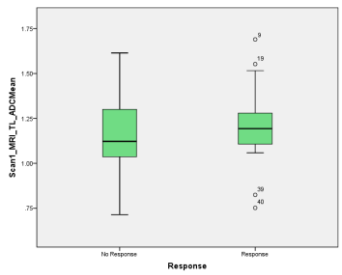
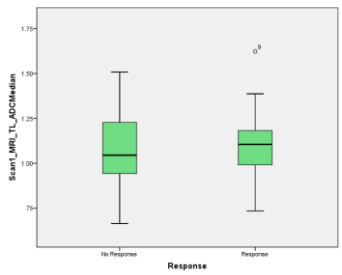
	ADC_{mean}	ADC_{median}
	Mean ($x10^{-3}$ mm^2s^{-1}) (95% CI) [min-max] {median; SD} |Skewness; Kurtosis|	
All lesions n=92, 20 patients	**1.16 (1.12-1.2)** [0.71-1.69] {1.13; 0.19} |0.32; 0.47|	**1.08 (1.04-1.12)** [0.67-1.62] {1.06; 0.19} |0.36; 0.02|
Box Plots All Lesions		
Breast cancer n=22	**1.27 (1.2-1.35)** [1-1.69] {1.28; 0.17} |0.62; 0.1|	**1.19 (1.12-1.27)** [0.95-1.62] {1.18; 0.18} |0.66; -0.04|
Prostate cancer n=70	**1.12 (1.08-1.17)** [0.71-1.61] {1.12; 0.18} |0.3; 0.68|	**1.05 (1.01-1.09)** [0.67-1.51] {1.03; 0.18} |0.35; -0.02|
Box Plots Tumour Groups		
p-value	0.001	0.01
No Response n=59 (B=10; p=49)	**1.16 (1.12-1.21)** [0.71-1.61] {1.12; 0.18} |0.45; 0.02|	**1.09 (1.04-1.14)** [0.67-1.51] {1.05; 0.19} |0.31; -0.55|
Response n=23 (B=12; p=60)	**1.2 (1.11-1.29)** [0.75-1.69] {1.19; 0.21} |0.22; 1.21|	**1.11 (1.03-1.2)** [0.74-1.62] {1.11; 0.19} |0.67; 1.35|
Box Plots Response groups		
p-value	0.363	0.601

Table 9-1: DW-MRI Per-lesion Analysis; Baseline ADC Parameters; Descriptive statistics, comparison of tumour groups (Mann-Whitney U tests) and comparison of response groups (Mann-Whitney U tests) (ADC expressed as 10^{-3} mm^2/s)

ADC_{mean} values are significantly larger than those of ADC_{median} (paired samples t-test: where $t(91)=10.837$, $p=4.6x10^{-18}$). The actual difference in the median of the distribution is only $0.08x10^{-3}$ mm^2/s.

Both ADC parameters are significantly lower for the prostate cancer lesions (independent samples t-tests: ADC_{mean}, where $t(90)=3.434$, $p=0.001$; ADC_{median} - $t(90)=3.353$, $p=0.01$). This

suggests a difference in the underlying bone metastasis physiology between the tumour types. Prostate cancer metastases classically are more sclerotic than lytic, and breast cancer bone metastases have a more mixed reaction. This lower $ADC_{mean/median}$ is likely to represent the increased sclerotic change within prostate cancer metastases.

There is no demonstrable difference in the distribution of ADC_{mean} or ADC_{median} between the response groups (independent samples t-tests: p=0.363 and p=0.601 respectively).

Tumour subgroup analysis demonstrates no difference in the distribution between the response groups for the prostate cancer patients ($ADC_{mean/median}$ p=0.871/0.782 respectively), or breast cancer patients ($ADC_{mean/median}$ p=1.00 /0.628 respectively).

9.2.2 tDV

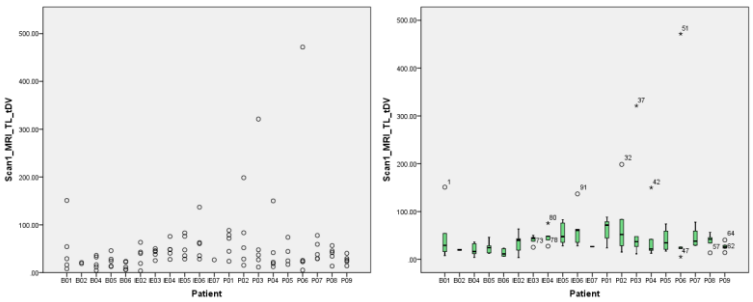


Figure 9-2: DW-MRI Per-lesion Analysis; Baseline tDV (cm3); Left – all lesions for all patients; right – tDV distributions for all patients

These whole-population plots (Fig.9-2) show a wide range of metastatic tumour volumes were identified. Among the prostate cancer patients there were 3 significantly larger metastases included, larger than 200cm^3. The population statistics are recorded in tabular and graphic form on the next page (Tbl.9-2).

Most lesions included were under 100cm^3, but several outliers can be noted with much larger volumes; these were prostate cancer patients. A statistically significant difference exists in the VOI volumes recorded from breast and prostate cancer metastases in this study (Mann-Whitney U test p=<0.0001). The breast cancer lesions were relatively more consistent in size.

These studies included breast and prostate cancer patients embarking on further lines of treatment; perhaps the prostate cancer patients had more significant disease burden at the time of inclusion in the study relative to the breast cancer patients.

The lesions which subsequently were classified as being in patients with a treatment response had a statistically lower baseline tDV (Mann-Whitney U test, p=0.002).

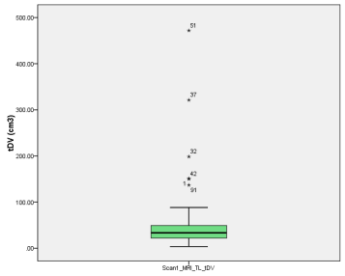
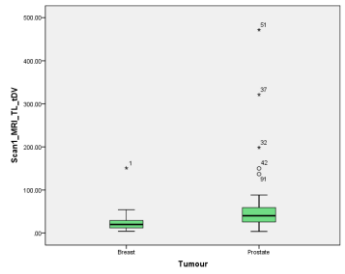
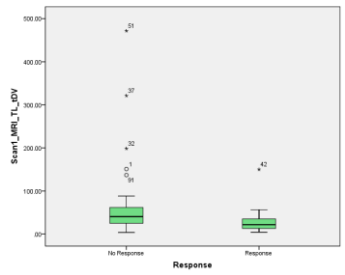
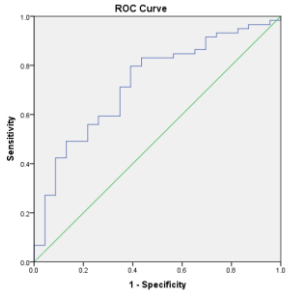
	Baseline tDV
	Mean (cm^3) (95% CI) [min-max] {median; SD} |Skewness; Kurtosis|
All lesions **n=92, 20 patients**	**48.67 (35.71-61.62)** [3.79-471.77] {33.5; 62.55} |4.64; 26.35|
Box Plot **All lesions**	
Breast cancer (B) **n=22**	**27.11 (13.59-40.63)** [4.46-151.03] {20.08; 30.5} |3.47; 13.93|
Prostate cancer (P) **n=70**	**55.44 (39.12-71.76)** [3.79-471.77] {40.31; 68.43} |4.38; 22.56|
Box Plot **Tumour Groups**	
p-value	<0.0001
No Response **n=59** **(B=10; p=49)**	**58.19 (38.89-77.49)** [3.79-471.77] {40.61; 74.07} |4.02; 18.83|
Response **n=23** **(B=12; p=60)**	**29.42 (16.58-42.26)** [4.46-150.11] {21.88; 29.69} |3.28; 13.02|
Box Plot **Response Groups**	
p-value	p=0.002

Table 9-2: DW-MRI Per-lesion Analysis; Baseline tDV; Descriptive statistics, comparison of tumour groups (Mann-Whitney U tests) and comparison of response groups (Mann-Whitney U tests)

The ROC analysis (Fig.9-3) for prediction of progression demonstrates an AUC of 0.726 (p=0.002) suggesting baseline tDV is an accurate predictor of PD.



AUC 0.726, p=0.002

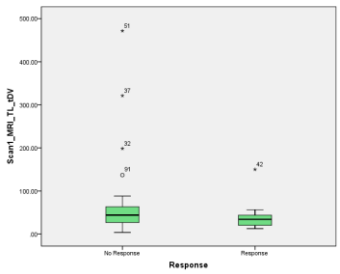
Figure 9-3: DW-MRI Per-lesion Analysis; ROC curve for prediction of 24-week response sub-group with baseline tDV; all patients

tDV predicts for progression in these patients with a sensitivity of 81% and specificity of 61% (for lesions greater than 23.76cm^3)(p=0.002). The prostate cancer lesions were larger than the breast lesions, and tDV may not be independent of the tumour group.

There is no statistically significant difference in tDV between the response groups for the prostate cancer patients (Mann-Whitney U test: p=0.178). However, the box-plot (Fig.9-4) of the prostate cancer data shows the patients with the largest lesions had PD by 24-weeks.

tDV

(prostate only)



p=0.178

Figure 9-4: DW-MRI Per-lesion Analysis; ROC curve for prediction of 24-week response sub-group with baseline tDV; prostate patients only

9.2.3 Heterogeneity Parameters

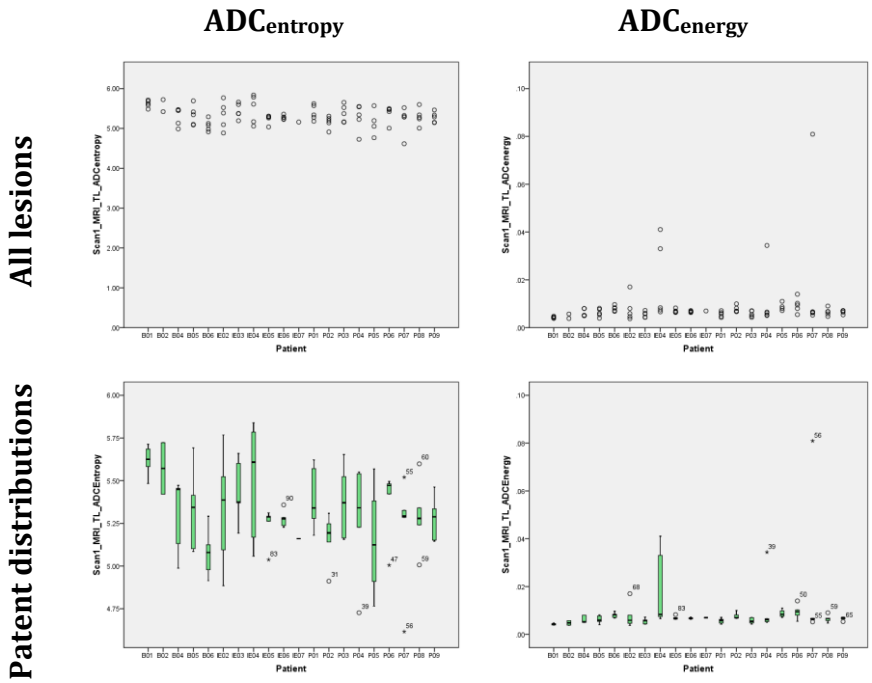


Figure 9-5: DW-MRI Per-lesion Analysis; Baseline ADC Heterogeneity Parameters ($ADC_{entropy}$-left, ADC_{energy}-right); TOP – all lesions for all patients; BOTTOM – ADC distributions for all patients (ADC expressed as 10-3 mm2/s)

The population data is recorded in Table 9-3. Figure 9-5 demonstrates the inter-lesion heterogeneity and the differences between the patients.

No difference in the $ADC_{entropy}$ distributions is shown between the tumour subgroups (independent samples t-test: where t(90)=0.768, p=0.445). The prostate lesions showed a trend towards a higher ADC_{energy} (Mann-Whitney U test: p=0.067) than the breast metastases.

The prostate lesion data had several outliers, all with higher ADC_{energy} measurements. This, suggests the prostate cancer lesions were generally more uniform (with a higher ADC_{energy}) than the breast cancer lesions. This may be reflecting the differing physiologies of the bone metastases; breast cancers tend to have bone lesions with a mix of osteolytic and osteoblastic changes, whereas prostate bone metastases tend to be much more osteoblastic, perhaps causing more ADC voxel distribution uniformity.

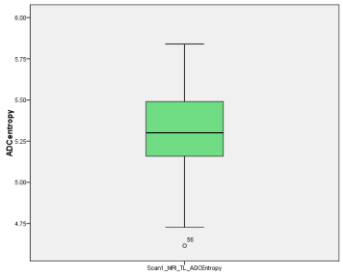
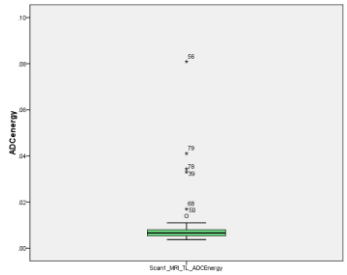
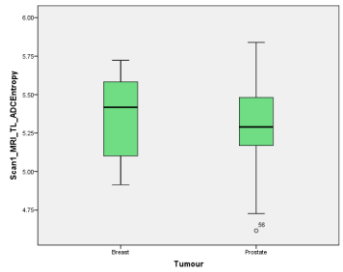
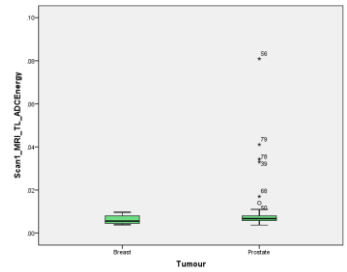
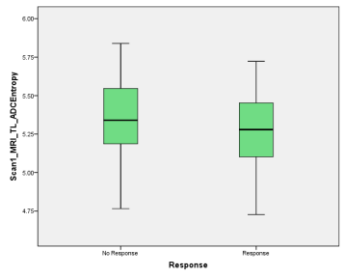
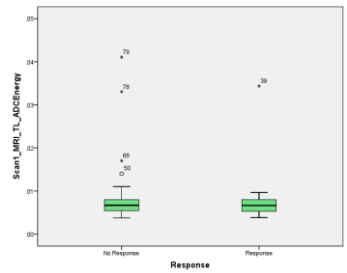
	Baseline $ADC_{entropy}$	Baseline ADC_{energy}
	Mean (95% CI) [min-max] {median; SD} |Skewness; Kurtosis|	
All lesions n=92, 20 patients	**5.32 (5.27-5.37)** [4.62-5.84] {5.3; 0.25} |-0.24; 0.03|	**0.0085 (0.0065-0.0104)** [0.0037-0.0809] {0.0066; 0.0095} |5.823; 39.2039|
Box Plots All lesions		
Breast cancer (B) n=22	**5.35 (5.24-5.47)** [4.91-5.72] {5.42; 0.26} |-0.15; -1.32|	**0.0061 (0.0053-0.0069)** [0.0038-0.0097] {0.0056; 0.0018} |0.3774; -1.1955|
Prostate cancer (P) n=70	**5.31 (5.25-5.36)** [4.62-5.84] {5.29; 0.24} |-0.32; 0.58|	**0.0092 (0.0066-0.0118)** [0.0037-0.0809] {0.0067; 0.0108} |5.1237; 30.0079|
Box Plots Tumour groups		
p-value	0.768	0.067
No Response n=59 (B=10; p=49)	**5.35 (5.29-5.41)** [4.77-5.84] {5.34; 0.24} |-0.05; -0.49|	**0.0078 (0.0062-0.0094)** [0.0037-0.0411] {0.0067; 0.006} |4.2927; 20.2816|
Response n=23 (B=12; p=60)	**5.26 (5.16-5.37)** [4.73-5.72] {5.28; 0.25} |-0.2; -0.43|	**0.0078 (0.0052-0.0103)** [0.0038-0.0344] {0.0066; 0.006} |4.3218; 19.8557|
Box Plots Response groups		
p-value	0.141	0.971

Table 9-3: DW-MRI Per-lesion Analysis; Baseline ADC Heterogeneity Parameters; Descriptive statistics, comparison of tumour groups (Mann-Whitney U tests) and comparison of response groups (Mann-Whitney U tests)

Patients with a treatment response had a lower baseline $ADC_{entropy}$, but similar ADC_{energy} compared to those in patients with PD, but these differences are not statistically significant difference ($ADC_{entropy}$ tested with independent samples t-test, where t(80)=1.486, p=0.141; ADC_{energy} tested with a non-parametric Mann Whitney U test, p=0.971). The tumour sub-group analyses can be reviewed in Figure 9-6 (data not included for brevity).

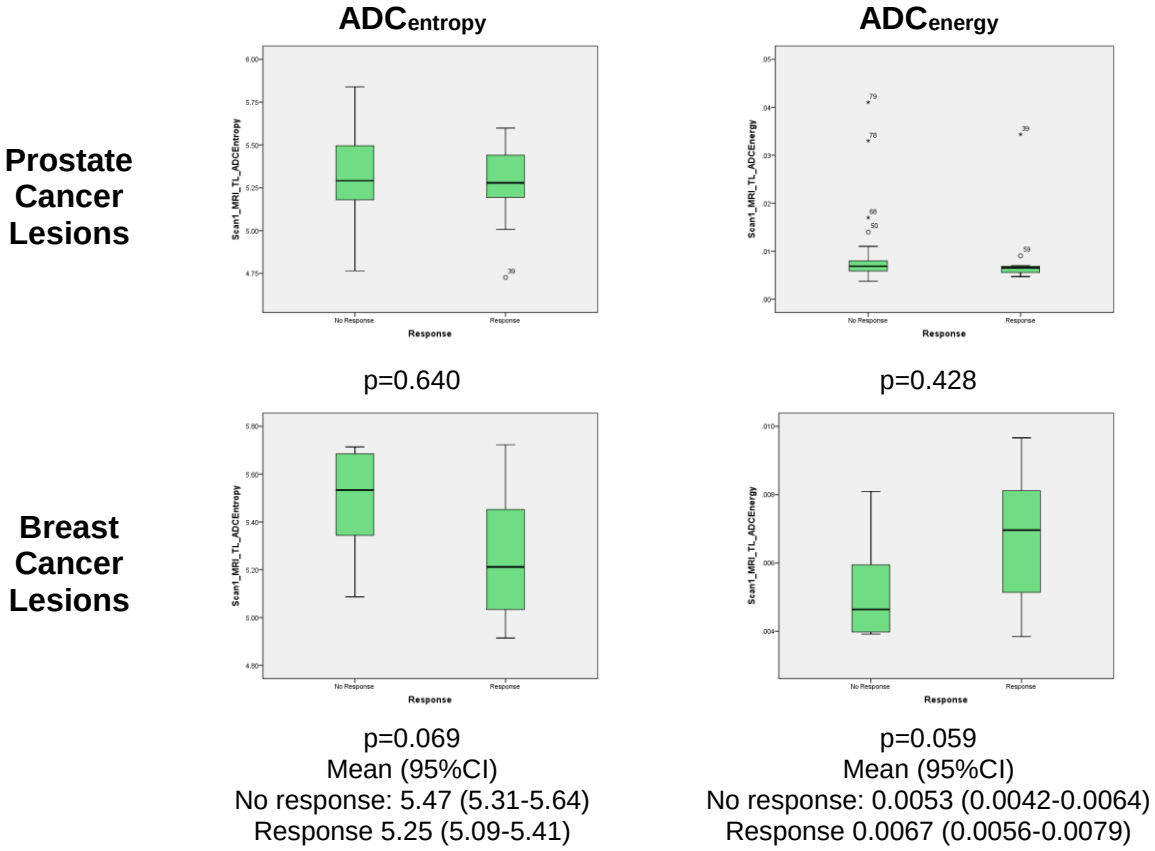


Figure 9-6: DW-MRI Per-lesion Analysis; Baseline Scan ADC Heterogeneity Parameters; Response Assessment; Tumour groups separately (comparison with Mann-Whitney U tests).

A lower baseline $ADC_{entropy}$ and higher ADC_{energy} (a more homogeneous picture) how a statistical trend for prediction of a treatment response (p=0.069 and 0.059 respectively).

9.2.4 Baseline Parameters Correlations

Table 9-4 summarises pairwise correlation analysis between the baseline scan parameters (analysed using Spearman Rank correlation tests).

	ADC_{mean}	ADC_{median}	tDV	$ADC_{entropy}$	ADC_{energy}
ADC_{mean}	1	-	-	-	-
ADC_{median}	0.93 (<0.001)	1	-	-	-
tDV	-0.09 (0.3864)	-0.11 (0.3085)	1	-	-
$ADC_{entropy}$	0.36 (0.0004)	0.35 (0.0006)	0.24 (0.0203)	1	-
ADC_{energy}	-0.33 (0.0013)	-0.31 (0.0025)	-0.18 (0.0851)	-0.78 (<0.001)	1

Table 9-4: DW-MRI Per-lesion Analysis; Baseline Parameters; Correlations (and p-values) using Spearman Rank analysis; All lesions

A strong correlation is identified between ADC_{mean} and ADC_{median}, as would be anticipated as they are describing the average of the same VOIs. There is no correlation between tDV and the other parameters. There is a only a weak positive correlation between $ADC_{entropy}$ and the ADC parameters, and a weak negative correlation between ADC_{energy} and the ADC parameters. The ADC heterogeneity parameters have a close negative correlation (correlation coefficient -0.78, p<0.001).

9.3 Second Scan

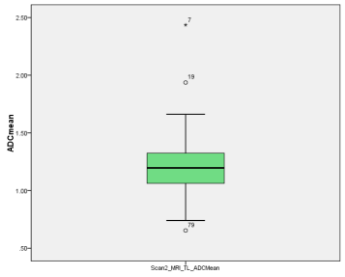
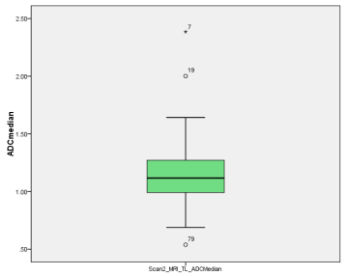
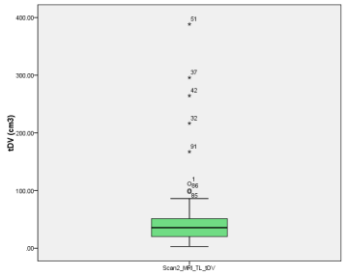
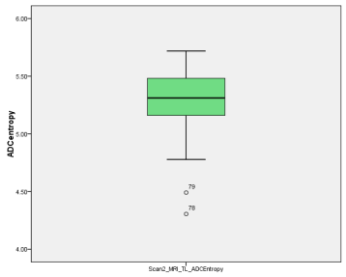
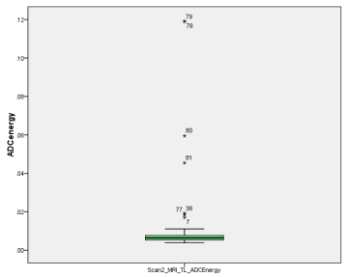
	Mean ($x10^{-3}$ mm^2s^{-1}) (95% CI) **[min-max]** **{median; SD} |Skewness; Kurtosis|**	**Box-Plot**
ADC_{mean} **10^{-3} mm^2/s**	**1.21 (1.16-1.27)** [0.65-2.44] {1.2; 0.25} |1.47; 6.01|	
ADC_{median} **10^{-3} mm^2/s**	**1.15 (1.09-1.2)** [0.54-2.39] {1.12; 0.27} |1.47; 5.37	
tDV **cm^3**	**50.68 (37.66-63.69)** [3.09-388.51] {35.75; 61.05} |3.57; 14.62|	
$ADC_{entropy}$	**5.3 (5.25-5.35)** [4.31-5.72] {5.31; 0.24} |-1.08; 3.01|	
ADC_{energy}	**0.0104 (0.0065-0.0143)** [0.0038-0.1191] {0.0064; 0.0183} |5.2485; 28.4132|	

Table 9-5: DW-MRI Per-lesion Analysis; Second Scan ADC Parameters; Descriptive statistics (ADC expressed as 10-3 mm^2/s)

9.4 % Change Between Scans

9.4.1 %ΔADC Parameters

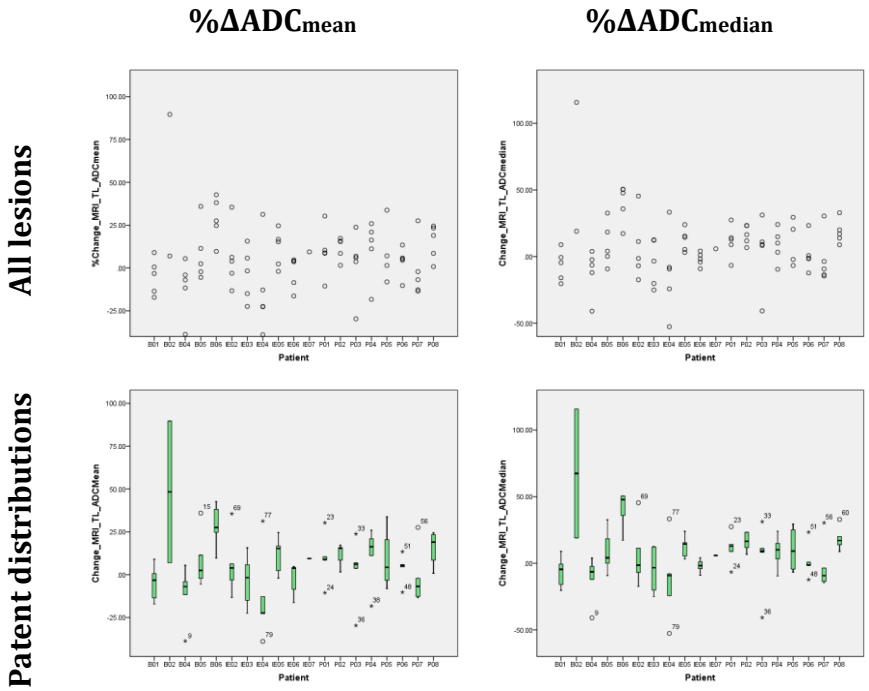


Figure 9-7: DW-MRI Per-lesion Analysis; %ΔADC Parameters (%ΔADC$_{mean}$-left, %ΔADC$_{median}$-right); TOP – all lesions for all patients; BOTTOM – ADC distributions for all patients

A wide range of percentage changes in the ADC parameters can be seen (Fig.9-7). Inter-lesion heterogeneity within individual patients is evident; within a single patient some lesions have an increased ADC, and others a decrease, suggesting a differential treatment effect. The population statistics are recorded in Table 9-6.

There is no demonstrable statistical difference in the distributions of the %Δ ADC parameters between the tumour groups (Mann Whitney U tests: %ΔADC$_{mean}$ p=0.838; %ΔADC$_{median}$, p=0.598).

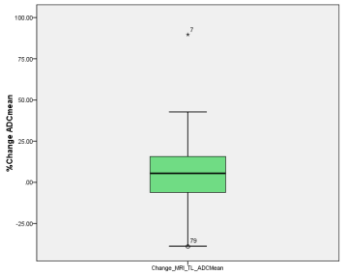
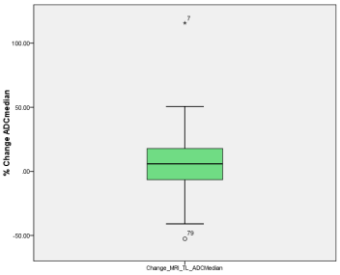
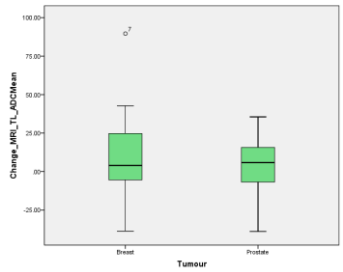
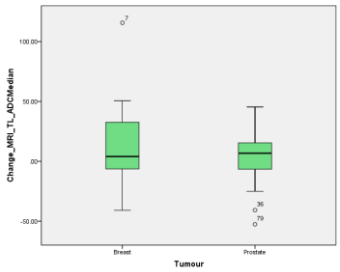
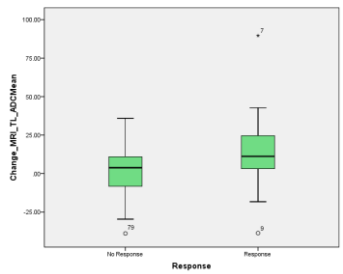
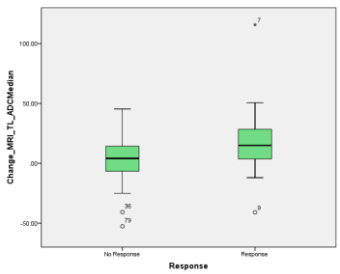
	%ΔADC$_{mean}$	%ΔADC$_{median}$
	% change (95% CI) [min-max] {median; SD} |Skewness; Kurtosis|	
All lesions n=82, 19 patients	**5.71 (1.63-9.78)** [-38.93-89.65] {5.47; 19.12} |0.79; 3.46|	**7.24 (2.4-12.07)** [-52.63-115.74] {5.84; 22.68} |1.08; 5.5|
Box Plots All lesions		
Breast cancer (B) n=22	**9.16 (-2.59-20.91)** [-38.72-89.65] {3.97; 26.5} |1.27; 3.08|	**13.29 (-1.38-27.96)** [-40.97-115.74] {4.01; 33.09} |1.39; 3.26|
Prostate cancer (P) n=65	**4.54 (0.58-8.5)** [-38.93-35.52] {5.81; 15.98} |-0.33; -0.08|	**5.19 (0.79-9.59)** [-52.63-45.42] {6.72; 17.75} |-0.57; 1.11|
Box Plots Tumour groups		
p-value	0.838	0.598
No Response n=59 (B-10; P-49)	**3.03 (-1.09-7.16)** [-38.93-35.93] {4.16; 15.82} |-0.11; 0.26|	**3.73 (-0.97-8.43)** [-52.63-45.42] {4.09; 18.04} |-0.47; 1.03|
Response n=23 (B-12; P-11)	**14.14 (3.44-24.84)** [-38.72-89.65] {11.18; 24.74} |0.85; 3.45|	**18.27 (5.23-31.31)** [-40.97-115.74] {14.93; 30.16} |1.34; 4.32|
Box Plots Response Groups		
p-value	0.018	0.024

Table 9-6: DW-MRI Per-lesion Analysis; %Δ ADC Parameters; Descriptive statistics, comparison of tumour groups (Mann-Whitney U tests) and comparison of response groups (Mann-Whitney U tests)

There has been a more positive change in ADC parameters between scans in the lesions from patients with a treatment response. These differences are statistically significant (%ΔADC$_{mean}$ data compared with an independent samples t-test: t(80)=-2.415, p=0.018; %ΔADC$_{median}$ data

compared with an independent samples Mann Whitney U test: p=0.024). Tumour subgroup analysis can be seen in Figure 9-8.

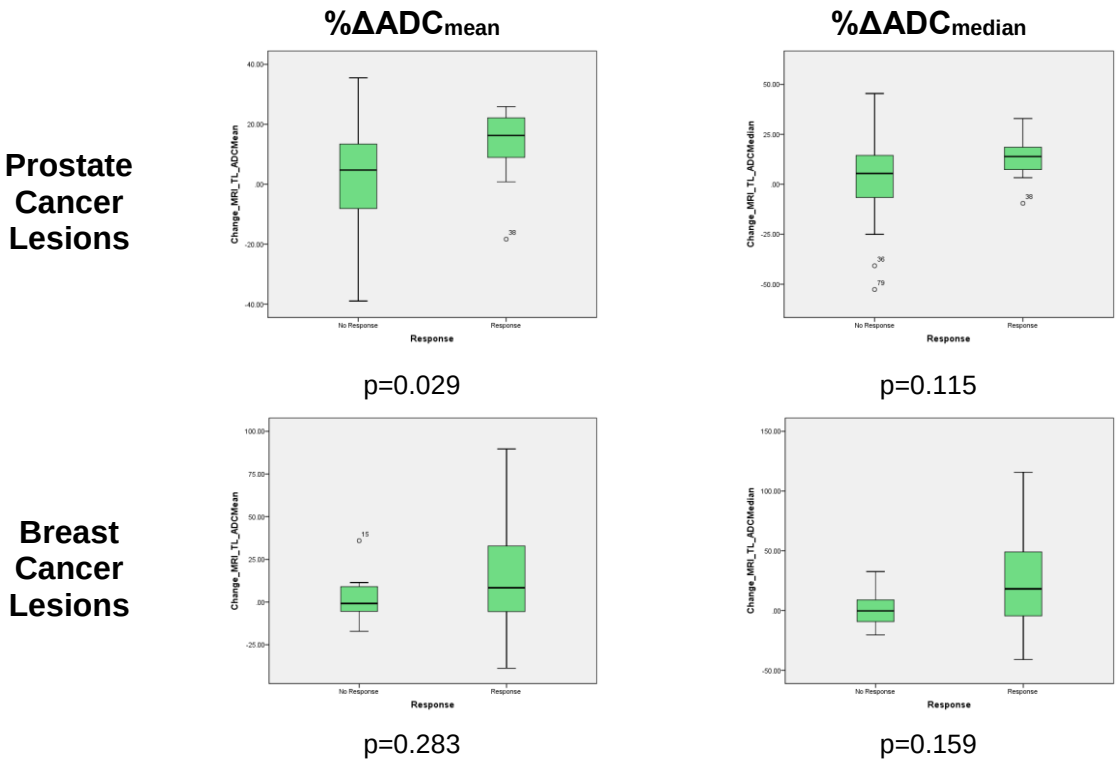


Figure 9-8: DW-MRI Per-lesion Analysis; %ΔADC Parameters; Response Assessment; Tumour groups separately

All the box-plots suggest higher ADCs in the response group, but only with ADC_{mean} for the prostate cancer lesions is statistical significant met (p=0.029).

ROC analysis shows a treatment response can be predicted for the prostate cancer lesions with a sensitivity of 81.8% and specificity of 67.3% with an increase in ADC_{mean} >8.53% between scans (p=0.029) (Fig.9-9).

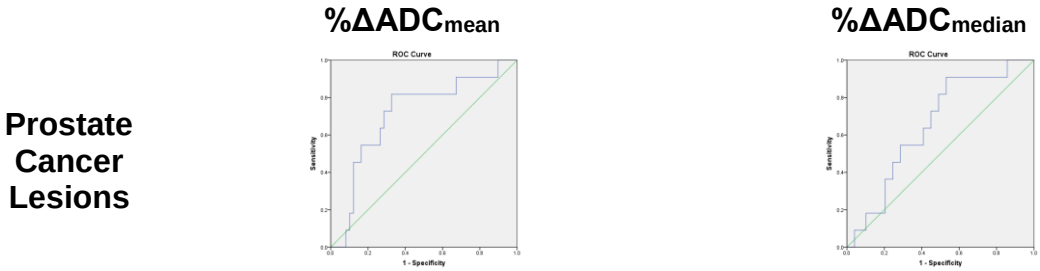


Figure 9-9: DW-MRI Per-lesion Analysis; %ΔΔADC Parameters; ROC analysis for predicting a treatment response; Prostate cancer lesions only

The individual lesion data can be used to construct waterfall plots (Figure 9-10). Despite the clear and demonstrable difference at a population level, the individual lesion data is more complex with a heterogeneous response. There are lesions showing a decrease in patients with a treatment response, and lesions showing less restriction in patients who have clinically PD.

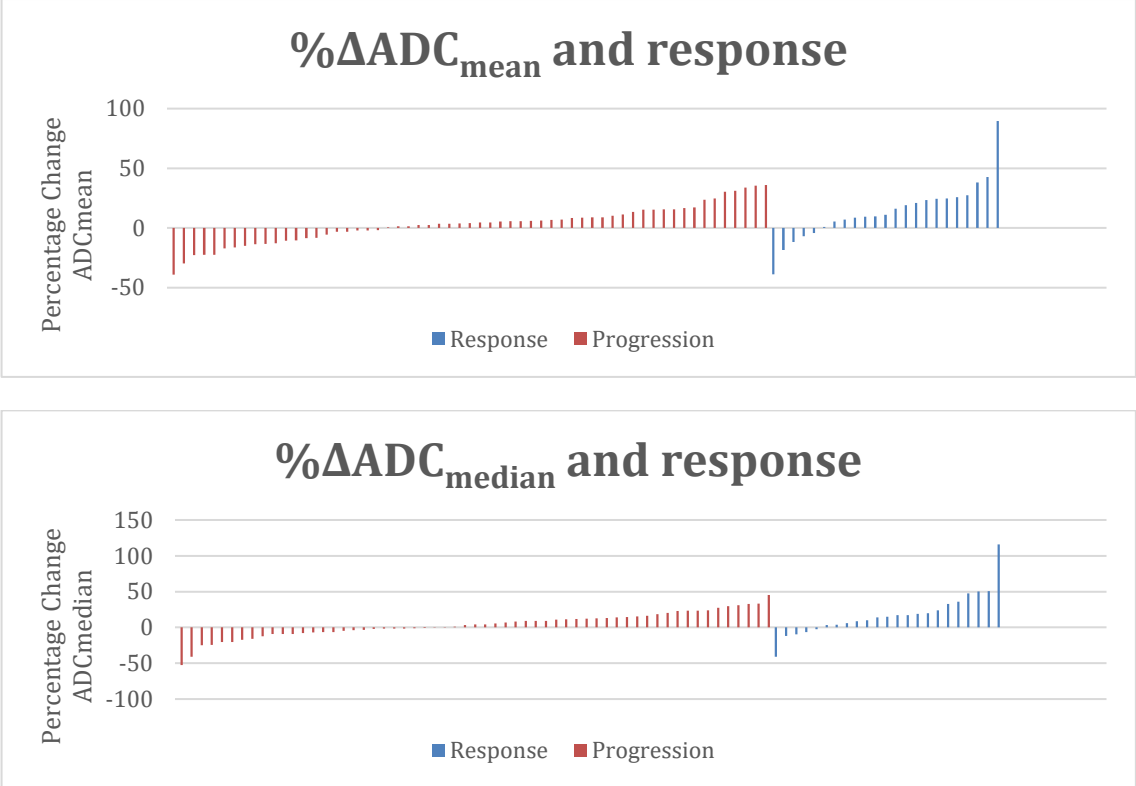


Figure 9-10: DW-MRI Per-lesion Analysis; Waterfall plots for %ΔADC Parameters; Response analysis

9.4.2 %Δ tDV

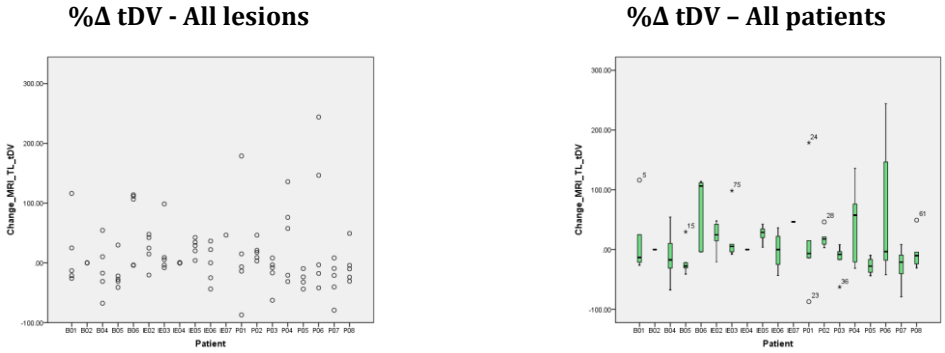


Figure 9-11: DW-MRI Per-lesion Analysis; %ΔtDV

The second scan tDV whole-population statistics suggested there was little difference with baseline imaging (there was no demonstrable statistical significance). Figure 9-11 shows the population analysis masked the underlying individual per-lesion changes. There is significant inter-lesion heterogeneity of response between lesions within the same patient; some lesions were larger on the second scan and others smaller. These changes will be compared with clinical response in a later section. The population statistics for %ΔtDV are recorded in Table 9-7.

On a study population level, there is no demonstrable difference in the %ΔtDV between the prostate cancer and breast cancer metastases (Mann Whitney U test: p=0.639).

There is no statistically significant difference in the distribution of %ΔtDV between the response groups (Mann Whitney U test: p=0.539).

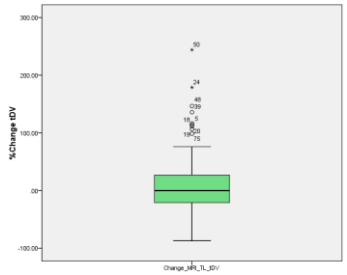
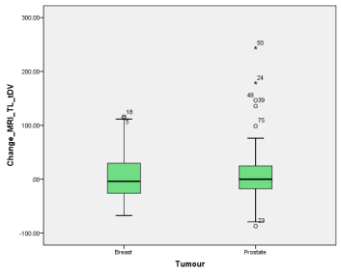
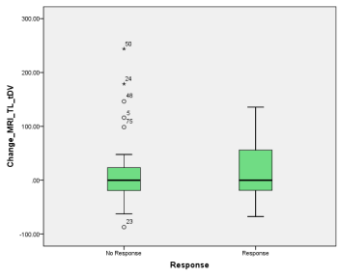
	%ΔtDV **Mean (95% CI)** [min-max] {median; SD} |Skewness; Kurtosis|
All lesions n=82, 19 patients	**11.7 (0.09-23.32)** [-87.06-243.99] {0; 54.5} |1.66; 4.06|
Box Plots All lesions	
Breast cancer (B) n=22	**11.96 (-12.28-36.19)** [-67.53-116.26] {-3.87; 54.65} |0.98; -0.13|
Prostate cancer (P) n=65	**11.62 (-1.98-25.22)** [-87.06-243.99] {0; 54.88} |1.91; 5.56|
Box Plots Tumour Groups	
p-value	0.639
No Response n=59 (B-10; P-49)	**10.87 (-3.28-25.03)** [-87.06-243.99] {0; 54.31} |2.17; 6.62|
Response n=23 (B-12; P-11)	**22.53 (-1.72-46.78)** [-67.53-135.79] {0; 56.07} |0.63; -0.66|
Box Plots Response Groups	
p-value	0.539

Table 9-7: DW-MRI Per-lesion Analysis; %Δ tDV; Descriptive statistics, comparison of tumour groups (Mann-Whitney U tests) and comparison of response groups (Mann-Whitney U tests) (ADC expressed as 10^{-3} mm^2/s)

The box-plots below (Fig.9-12) record the response analysis for the tumour groups considered separately.

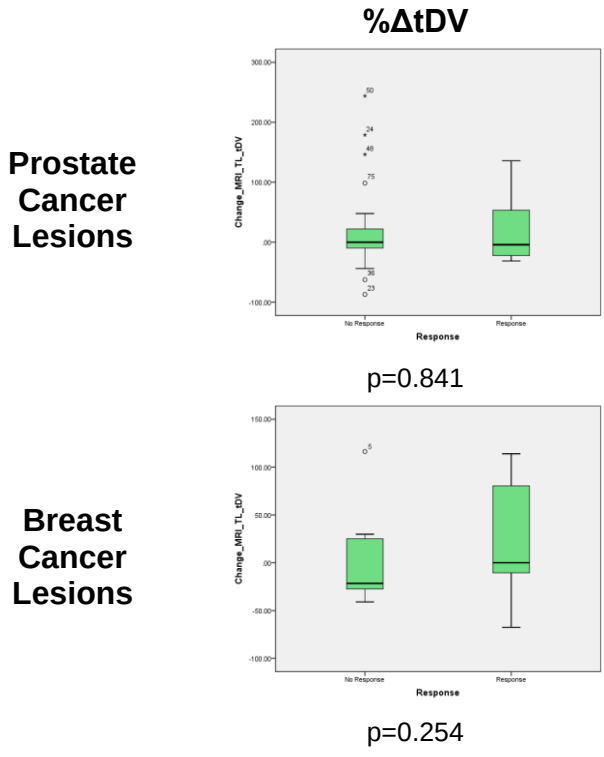


Figure 9-12: DW-MRI Per-lesion Analysis; Response Analysis using %ΔtDV; Tumour subgroup analyses (comparison with Mann-Whitney U tests)

Although the distribution of %Δ tDV between the response groups does not reach statistical significance for either tumour subgroup, it is striking to identify an increase in tDV in patients with a clinical response. This is an unexpected finding. Similarly, there are lesions with a decrease in tDV from patients who overall had no treatment response. There is considerable inter-lesion heterogeneity of response. This heterogeneity in response using this parameter can be more clearly seen with a waterfall plot (Fig.9-13). There were some significant increases in measured tDV in both response groups, with some measured lesions doubling in size.

There may indeed be a truly heterogeneous response within patients with some lesions increasing in size and others decreasing. The clinical response assessment considers other clinical and radiological parameters (as per the PCWG2 guidelines), and the individual lesion changes may not impact the clinical response assessment i.e. they are not clinically dominant lesions. Alternatively, the tDV may not be measuring the true tumour volume. DW-MRI is assessing the water diffusion of the tumour cells but also the bone stroma in which the tumour deposits are growing. Sclerosis causes restricted diffusion, and sclerosis is a known treatment response of healing bone metastases. The tDV is defined by the area of image restriction, and perhaps the peri-tumour bone undergoing sclerosis results in a larger defined tDV. It is not possible to separate the DW-MRI signal from tumour and bone stroma.

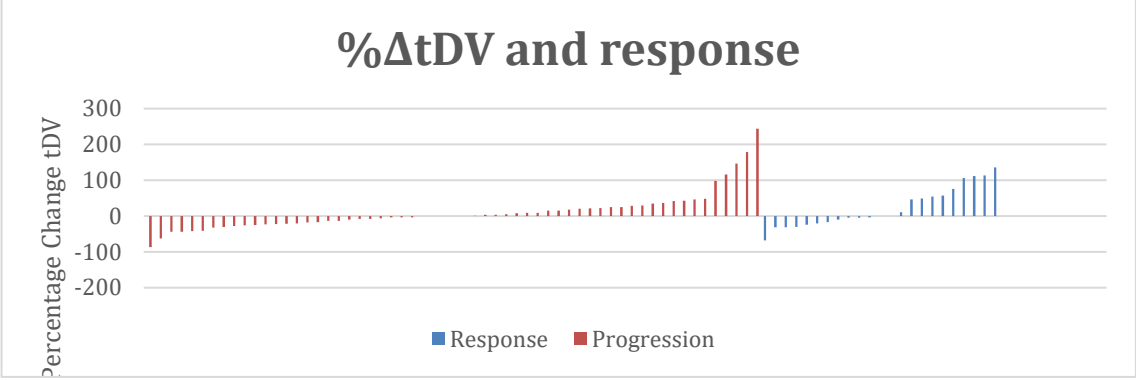


Figure 9-13: DW-MRI Per-lesion Analysis; Waterfall plot for %ΔtDV Parameters; Response analysis

9.4.3 %Δ ADC Heterogeneity Parameters

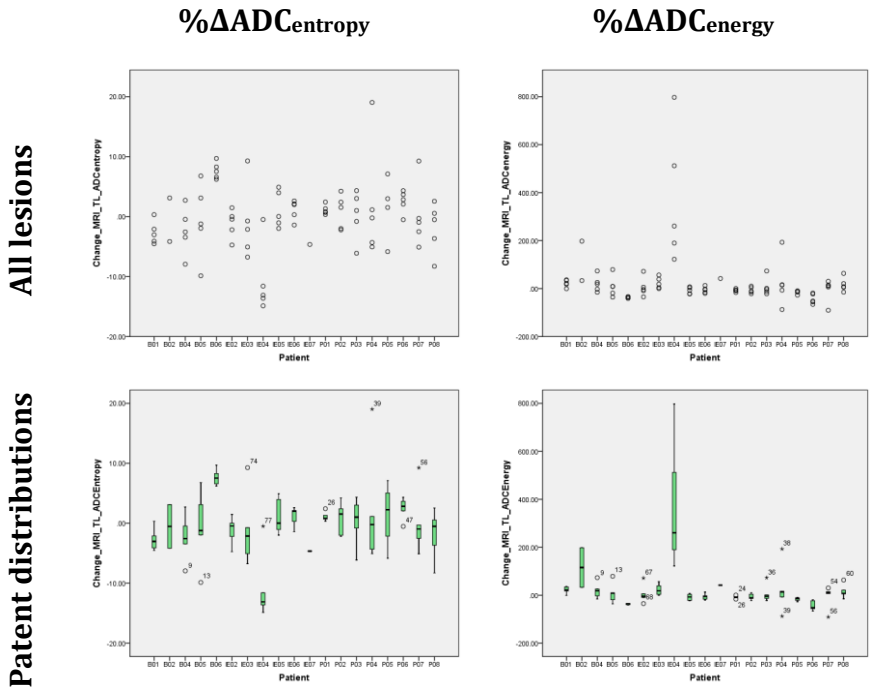


Figure 9-14: DW-MRI Per-lesion Analysis; %ΔADC Parameters (%ΔADC$_{mean}$-left, %ΔADC$_{median}$-right)

Like the other parameters discussed so far, the comparison of population statistics between baseline and second imaging did not show significant changes, but lesion based analysis shows a wide range of change in the ADC heterogeneity parameters, with inter-lesion variation between patients, and again within patients (Fig.9-14). There are contradictory changes in parameter within individual patients, with measured values increasing in some lesion and decreasing in others. The population statistics for change in heterogeneity parameters are recorded in Table 9-8.

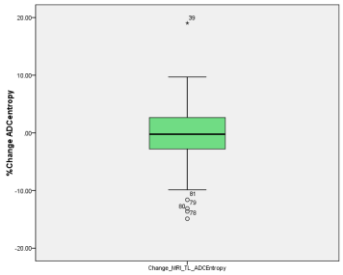
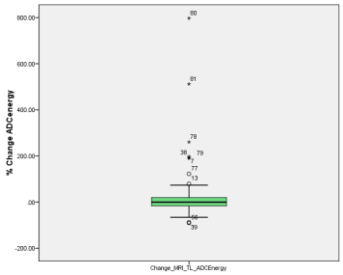
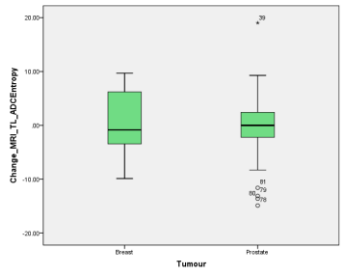
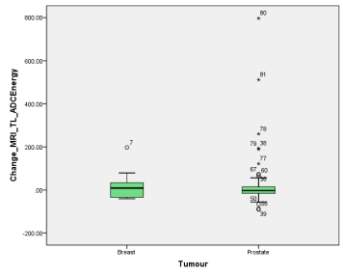
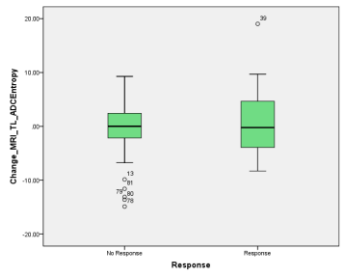
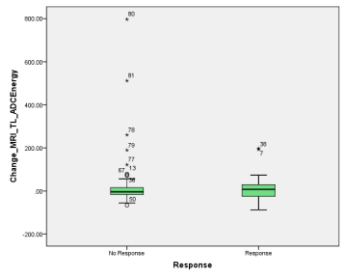
	$\%\Delta ADC_{entropy}$	$\%\Delta ADC_{energy}$
	% change (95% CI) [min-max] {median; SD} |Skewness; Kurtosis|	
All lesions n=82, 19 patients	**-0.21 (-1.35-0.94)** [-14.88-19.04] {-0.22; 5.37} |0.03; 1.87|	**24.95 (0.6-49.29)** [-90.47-797.16] {-0.03; 114.21} |4.77; 27.54|
Box Plots All lesions		
Breast cancer (B) n=22	**0.41 (-2-2.83)** [-9.86-9.7] {-0.84; 5.45} |0.08; -0.87|	**13.6 (-10.28-37.49)** [-40.75-197.7] {8.89; 53.87} |1.99; 5.72|
Prostate cancer (P) n=65	**-0.42 (-1.75-0.91)** [-14.88-19.04] {0.01; 5.37} |0.01; 2.91|	**28.78 (-3.06-60.63)** [-90.47-797.16] {-2.41; 128.52} |4.41; 22.53|
Box Plots Tumour Groups		
p-value	0.703	0.675
No Response n=59 (B-10; P-49)	**-0.69 (-1.97-0.59)** [-14.88-9.28] {0.02; 4.91} |-1.02; 1.4|	**30.46 (-3.87-64.78)** [-65.89-797.16] {-3.8; 131.71} |4.49; 22.46|
Response n=23 (B-12; P-11)	**0.96 (-1.82-3.75)** [-8.29-19.04] {-0.22; 6.45} |0.96; 1.29|	**17.44 (-11.5-46.38)** [-87.32-197.7] {7.63; 66.92} |1.58; 3.05|
Box Plots Response Groups		
p-value	0.631	0.617

Table 9-8: DW-MRI Per-lesion Analysis; %Δ ADC Heterogeneity Parameters; Descriptive statistics, comparison of tumour groups (Mann-Whitney U tests) and comparison of response groups (Mann-Whitney U tests) (ADC expressed as 10-3 mm^2/s)

At a whole-population level, there is no demonstrably statistically significant difference between tumour groups of the distribution of $\%\Delta ADC_{entropy}$ (Mann Whitney U test, p=0.703) or $\%\Delta ADC_{energy}$ (p=0.675).

There is no clear observable pattern in the distributions of the %ΔADC heterogeneity parameters between the two response groups (Mann-Whitney U tests: $\%\Delta ADC_{entropy}$ p=0.631; $\%\Delta ADC_{energy}$ p=0.617).

Tumour subgroup analysis can be seen below (Fig.9-15):

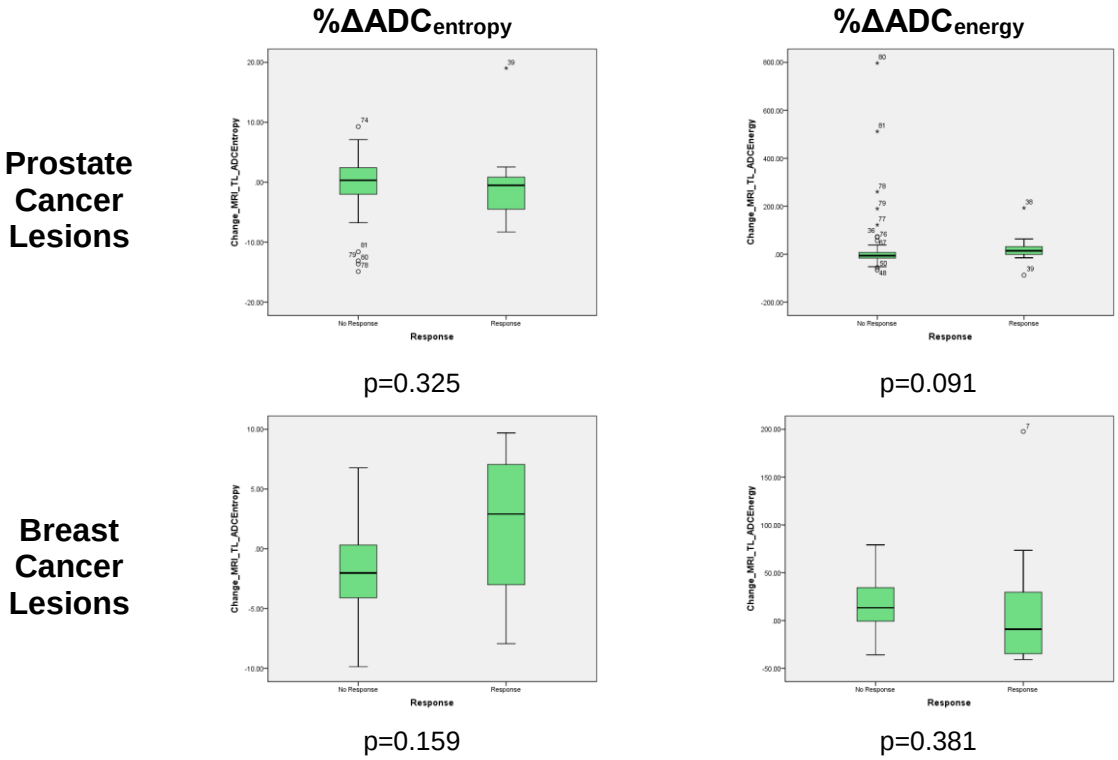


Figure 9-15: DW-MRI Per-lesion Analysis; %ΔADC Heterogeneity Parameters; Response Assessment; Tumour groups separately

There is no statistically significant difference identified using either %ΔADC heterogeneity parameter between the response groups, but there is perhaps a trend towards significance for $\%\Delta ADC_{energy}$ for the prostate cancer lesions (p=0.091) with the responding lesions tending towards a larger increase in ADC_{energy}, i.e. a move towards uniformity of the ADC voxel distributions within the VOIs.

As with the other parameters examined, there is considerable variation in how the ADC heterogeneity parameters responded to treatment in these target lesions. The waterfall plots (Fig.9-16) display the individual lesion changes separated into the 24-week response assessment groups.

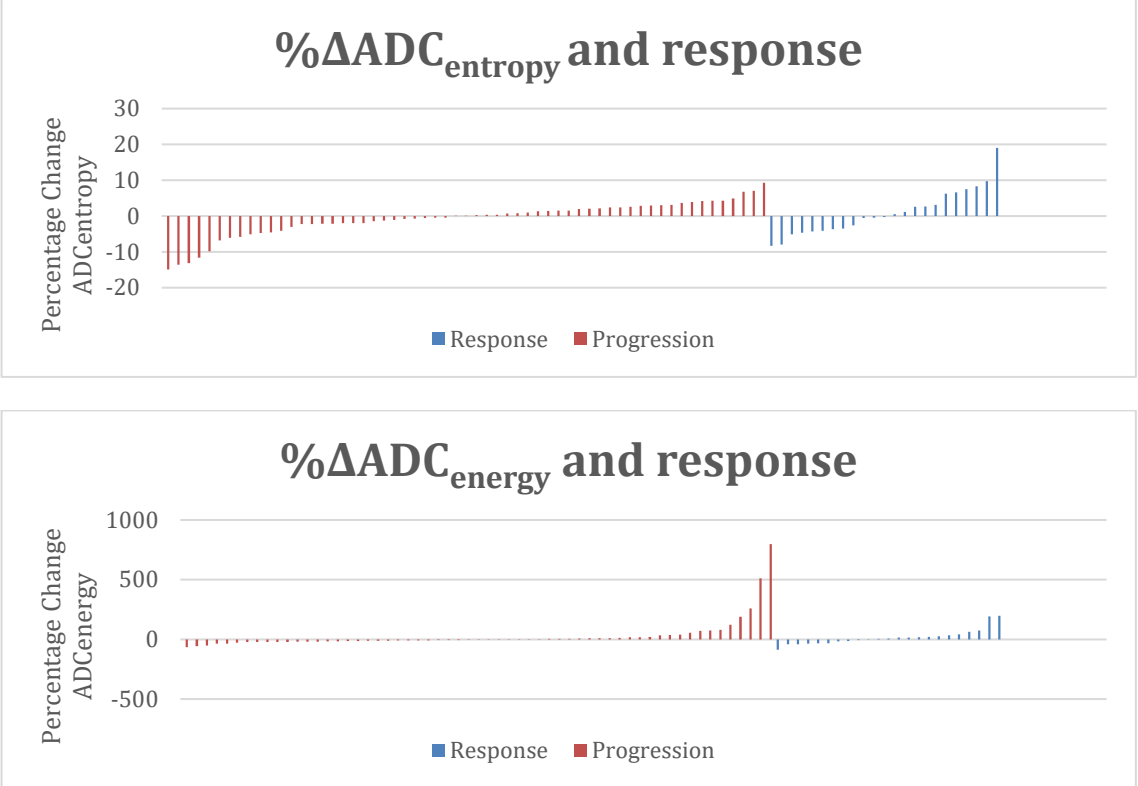


Figure 9-16: DW-MRI Per-lesion Analysis; Waterfall plots for %ΔADC Heterogeneity Parameters; Response analysis

9.4.4 %Change Parameters Correlations

Table 9-9 summarises pairwise correlation analysis between the %Δ parameters (analysed using Spearman Rank correlation tests).

	%ΔADC$_{mean}$	%ΔADC$_{median}$	%ΔtDV	%ΔADC$_{entropy}$	%ΔADC$_{energy}$
%ΔADC$_{mean}$	1	-	-	-	-
%ΔADC$_{median}$	0.95 (<0.001)	1	-	-	-
%ΔtDV	0.19 (0.075)	0.14 (0.2088)	1	-	-
%ΔADC$_{entropy}$	0.39 (0.0002)	0.37 (0.0005)	0.09 (0.3981)	1	-
%ΔADC$_{energy}$	-0.39 (0.0002)	-0.34 (0.0011)	-0.08 (0.4386)	-0.83 (<0.001)	1

Table 9-9: DW-MRI Per-lesion Analysis; Baseline Parameters; Correlations (and p-values) using Spearman Rank analysis; All lesions

A strong correlation is identified between %ΔADC$_{mean}$ and %ΔADC$_{median}$ parameters, as would be anticipated as they are describing the average of the same VOIs. There is no correlation between %ΔtDV and the other parameters. There is only a weak but significant positive correlation between %ΔADC$_{entropy}$ and the %ΔADC parameters, and a weak but significant negative correlation between %ΔADC$_{energy}$ and the %ΔADC parameters. The %ΔADC heterogeneity parameters have a close negative correlation (correlation coefficient -0.83, p<0.001).

9.5 Discussion

		ADC_{mean}	ADC_{median}	tDV	$ADC_{entropy}$	ADC_{energy}
Baseline Scan	Descriptive	Inter-lesion heterogeneity	Inter-lesion heterogeneity	Inter-lesion heterogeneity	Inter-lesion heterogeneity	Inter-lesion heterogeneity
	Tumour	Breast>Prostate p=0.001	Breast>Prostate p=0.01	Prostate>Breast P<0.0001	X	Prostate>Breast p=0.067
	Response	X	X	X	↓ Breast only p=0.069	↑ Breast only p=0.059
%Δ	Descriptive	Inter-lesion heterogeneity	Inter-lesion heterogeneity	Inter-lesion heterogeneity	Inter-lesion heterogeneity	Inter-lesion heterogeneity
	Tumour	X	X	X	X	X
	Response	↑ (Pca) p=0.029	↑ (Pca) p=0.115	X	X	↑ (Pca) p=0.091

Table 9-10: DW-MRI Per-lesion Analysis; All Response Analyses; Summary Table

Inter-lesion heterogeneity has been identified at the per-lesion level, irrespective of the quantification factor measured. Within individual patients there are bone metastases with notably different measurements, and the %Δ parameters following treatment do not provide a consistent direction or magnitude of response in individual patients. Inter-lesion heterogeneity and inconsistency of changes with treatment within individual patients has been described in bone metastases assessed with DW-MRI[152].

This inter-lesion heterogeneity within patients is clearly seen with the baseline ADC data. Despite the heterogeneity, patterns can be demonstrated. Prostate cancer lesions demonstrate more diffusion restriction at baseline (ADC_{mean} Breast 1.27 (+/-0.04), Prostate 1.12 (+/- 0.04), p=0.001, ADC_{median} Breast 1.19 (+/-0.07), Prostate 1.05 (+/-0.04), p=0.01). This probably reflects the difference in the physiology of the bone metastases; although all bone metastases cause both lytic and sclerotic changes within the bone, prostate cancer metastases generally result in a significantly osteoblastic sclerotic response whereas breast cancer can be more variable. Sclerotic bone can result in more restriction of water diffusion, measured as a lower ADC[164,212]. Tumours cause restriction of water diffusion because of higher cell density and tumour tissue disorganisation. However, imaging bone metastases includes measurement of the tumour as well as the changes within the surrounding bone.

None of the baseline parameters predicted for the 24-week reference standard treatment response. There is no evidence in the literature to suggest ADC measurements predict the ultimate clinical treatment response of bone metastases. An increasing body of evidence suggests a lower pre-treatment ADC in tumour tissue (not necessarily bone metastases) is associated with an increased response rate[145,213-219]. The correlation of ADC with the underlying histology is more complicated. A lower ADC of primary prostate cancer correlates with a higher Gleason grade, indicating a more aggressive and proliferative histology[220], and there is an association between the Gleason Grade and PFS[221]. Others have demonstrated no difference in the baseline ADC measurements of bone metastases from prostate cancer between responders and non-responders[152].

There is evidence suggesting the prognostic value of the volume of osseous disease in prostate cancer[222,223], but there is no evidence showing utility for predicting a treatment response. This study has found no association either.

There is more evidence in the literature of the possible utility of baseline ADC heterogeneity parameters. There is a correlation between heterogeneity and the prostate cancer Gleason Grade (increase in Gleason grade is associated with a higher $ADC_{entropy}$ and lower ADC_{energy}[172,176,177,224] – this suggests the more aggressive tumours (higher Gleason grade) have a more heterogeneous ADC voxel distribution. This might translate into a predictor of treatment response. In this study the ADC heterogeneity parameters have only suggested utility for predicting the treatment response for the breast cancer metastases. A treatment response was more likely for patients with a lower $ADC_{entropy}$ and higher ADC_{energy} (i.e. more homogeneous ADC voxel distributions); this finding was approaching statistical significance for ADC_{energy} ($p=0.067$).

Interpreting the literature is complicated by significant variations in the definition of a treatment response and the timing of the response assessment. Most studies use the RECIST criteria for determining a treatment response in a soft tissue tumour citing a partial response with at least a 30% reduction in the tumour volume[215,216,219,225], including a study of bone metastases, where soft tissue response as assessed with the RECIST criteria was used as a surrogate marker of treatment response[152]. This study, by Messiou et al., is most comparable to the analyses in this thesis, with MRI repeated at 12 weeks after initiation of chemotherapy for bone-predominant metastatic prostate cancer; 26 men were analysed, but there was no demonstrable difference in the ADC of the bone lesions before therapy when the group was dichotomised into the 12-week response group[152]. Treatment related changes in ADC quantification occur before any significant tumour volumetric change. An early response may not translate into a clinically relevant sustained response for the patient; identifying a gold-standard for response definition is difficult, particularly for bone metastases not amenable to size assessment.

This data shows a larger increase in the ADC parameters between scans predicts for a treatment response. Subgroup analysis suggests this relationship exists for both tumour types, but the differences only reach statistical significance for the prostate cancer lesions and with $\%\Delta ADC_{mean}$ ($p=0.029$).

A treatment response associated with a rise in measured ADC has been reported by several other groups in non-bone and non-prostate cancers[142,143,215,217-219,226-228]; an increase in ADC appears to be prognostic. This has also been demonstrated in malignancies involving the skeleton[58,148,157,229-231], and specifically prostate cancer bone metastases[151-153,232,233]. This finding in this study could therefore be anticipated. However, there do appear to be differences in both the measurements, and how the parameters change with treatment, depending on the tumour type and the specific pathophysiological processes within the peri-tumour bone environment.

Graham et al. (2014) examined a preclinical model of prostate cancer bone metastases, assessing the ADC changes following treatment with Cabozantinib. They also showed an increase in ADC is associated with a treatment response, attributing this change with increased

necrosis[232]. Increased necrosis and a corresponding rise in measured ADC has been demonstrated histologically in osteogenic sarcomas[148].

The inter-lesion heterogeneity of change in ADC parameters in individual patients is significant, not only with magnitude of parameter change, but also direction of change. This heterogeneous response was reported by Messiou et al., 2011 in their study of prostate cancer bone metastases, and is therefore directly relatable to this current study. This group attributed the heterogeneous response to varying effects of treatment on the composition of the bone marrow[152] – a different impact will result depending on the haematopoietic cell composition within the bone being quantified. Bone marrow is composed of fat cells and, in varying proportions, haematopoietically active marrow which contains more water than the adjacent less active marrow and the fat cells, and thus demonstrates more free movement of water and a higher ADC. An increase in haematopoietic cells has been shown to result in a corresponding decrease in the more diffusion restricting fat cells, and vice versa[152]. With PD the measured bone ADC may counterintuitively increase after replacement of tumour cells with normal fatty marrow; progression of an osteolytic component of disease will result in a higher measured ADC; the healing sclerotic reaction of a treatment response in bone can result in a fall in ADC[152]. Reishauer et al. also noted a heterogeneous response, and identified areas of ADC fall within 'responding' tumours with repeated imaging, suggesting an effect of post-treatment fibrosis, or repopulation of the tumour area by fatty tissue and/or red marrow, or indeed could represent areas of sclerosis suggesting healing bone[153]. An ADC response is subject to numerous conflicting variables, and this perhaps explains the heterogeneous changes seen in these patients.

Assessment of bone metastases is significantly complicated by the unavoidable influence of the peri-tumour bone architecture, the function of which is constantly impacted by the tumour, by the therapy, and by subsequent healing or damage.

The %Δ tDV in these patients has not demonstrated any utility for predicting treatment response. There is significant intra-patient heterogeneity of %ΔtDV. An increase in tDV of bone metastases from breast and prostate cancer is associated with non-response of the disease[154], although these changes in tDV were measured after an interval of 12-38 weeks (mean of 22 weeks). The 8-12-week interval between scans in this study might be too short to identify significant changes in the tDV. Many studies showing prognostic utility of DW-MRI have defined response by reduction in the size of soft tissue malignancies With this analysis of DW-MRI perhaps tumour and abnormal peri-tumour bone is included in the quantification, complicating quantification analysis.

The %Δ ADC Heterogeneity parameters have not demonstrated a prognostic utility in this study, for either tumour group, but the now familiar inter-lesion heterogeneity in individual patients is clearly demonstrated.

In summary, per-lesion analysis has clearly shown intra-patient inter-lesion heterogeneity, and the parameters from prostate cancer metastases are different from breast cancer; the sclerosis of prostate cancer bone metastases probably results in a lower ADC. A larger increase in ADC between scans, measured at 8-12 weeks, predicts the 24-week treatment response.

Chapter 10 DW-MRI Per-lesion Analysis - Survival assessment using DW-MRI quantification parameters of patients with bone metastases from breast or prostate cancer

The methods of this chapter and the statistical approaches for this section can be reviewed in Chapter 2.3.9.

10.1 Lesion Based Analysis – OS Analysis

10.1.1 Tumour Analysis

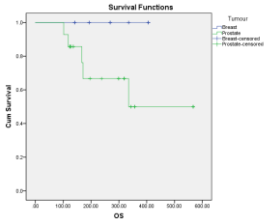


Figure 10-1: DW-MRI Per-lesion Analysis; KM OS Analysis; Tumour Groups

The breast cancer patients in this study had a significantly better OS than the prostate cancer patients – none of the breast cancer patient had died at the time of data analysis, but 5 of the prostate cancer patients had (log rank p=0.002, Fig.10-1)).

10.1.2 Baseline Scan

10.1.2.1 ADC Parameters

	ADC_{mean}	ADC_{median}
	Cox regression p-value [HR (95%CI)]	
n=92, 25 events	**0.46** [0.464 (0.06-3.561)]	**0.68** [0.651 (0.084-5.026)]

Table 10-1: DW-MRI Per-lesion Analysis; Baseline ADC Parameters; OS Cox regression analysis

Cox regression analysis (Tbl.10-1) does not demonstrate correlation between the $ADC_{mean/median}$ with the length of OS. The data can also be examined with KM analysis - these plots are displayed in Figure 10-2.

Lesions with a baseline ADC_{mean} greater than 1.197 or ADC_{median} greater than 1.14 have a statistically longer OS than those with lower values.

The baseline ADC parameters did not show utility in predicting the response for these patients, and an OS benefit was not anticipated. The baseline ADC parameters were significantly higher for the breast cancer patients, and this survival benefit could simply be selection of the lesions from breast cancer patients. KM analysis of the prostate patients confirms this statistically; a lower baseline ADC predicts for a longer OS (ADC_{mean} p=0.047; ADC_{median} p=0.005) (Fig.10-3). Baseline ADC parameters predict OS for prostate cancer lesions. Sclerosis in bone results in

more diffusion restriction and a lower ADC[234]. Increased sclerosis has been associated with a survival benefit[235].

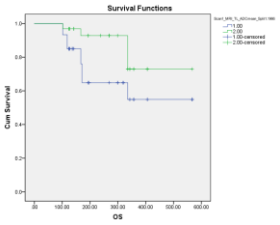
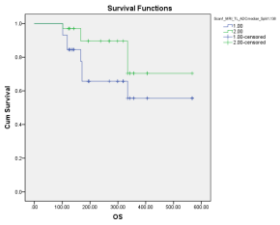
ADC_{mean}	ADC_{median}
	
Dichotomised at 1.1966 Low (≤) – Blue High (>) - Green	Dichotomised at 1.138 Low (≤) – Blue High (>) - Green
p=0.03	p=0.054

Figure 10-2: DW-MRI Per-lesion Analysis; KM OS Analysis; baseline ADC parameters; all patients

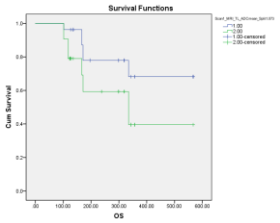
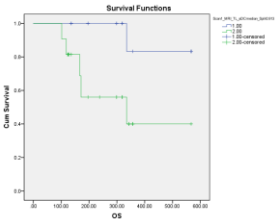
ADC_{mean} (prostate only)	ADC_{median} (prostate only)
	
Dichotomised at 1.073 Low (≤) – Blue High (>) - Green	Dichotomised at 0.913 Low (≤) – Blue High (>) - Green
p=0.047	p=0.005

Figure 10-3: DW-MRI Per-lesion Analysis; KM OS analysis; baseline ADC parameters; Prostate cancer patients only

10.1.2.2 tDV

	tDV
	Cox regression p-value [HR (95%CI)]
n=92, 25 deaths	**0.678** [1.001 (0.995-1.007)]

Table 10-2: DW-MRI Per-lesion Analysis; Baseline tDV; OS Cox regression analysis

Cox regression analysis (Tbl.10-2) of the baseline tDV data does not demonstrate a significant correlation between tDV and OS. There is a statistically significant separation of the two KM survival curves (Fig.10-4) (Log rank, p=0.044). Patients with smaller lesions at baseline had a significantly better OS.

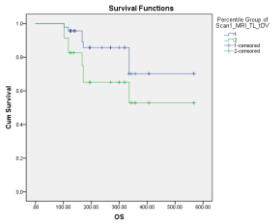
tDV

Dichotomised at 33.5 (median) Low (≤) – Blue High (>) - Green
p=0.044

Figure 10-4: DW-MRI Per-lesion Analysis; KM OS Analysis; baseline tDV; All Patients

However, the baseline tDV discriminated between the breast and prostate cancer lesions; prostate lesions were larger, and the prostate patients had a shorter OS. KM analysis with only the prostate cancer patients does not identify utility of baseline tDV for predicting OS (dichotomisation at 31.92cm^3, p=0.354). Analysis of just the breast cancer patients is not feasible as there were no deaths.

Therefore, tDV has not been shown to be an independent predictor of OS.

10.1.2.3 Heterogeneity Parameters

	$ADC_{entropy}$	**ADC_{energy}**
	Cox regression p-value [HR (95%CI)]	
n=92; 25 deaths	**0.675** [0.714 (0.147-3.457)]	**0.335** [0 (0-2.44E+25)]

Table 10-3: DW-MRI Per-lesion Analysis; Baseline ADC Heterogeneity Parameters; OS Cox regression analysis

Cox regression analysis (Tbl.10-3) does not indicate a significant association between the baseline ADC heterogeneity parameters and the OS. KM analysis produces the survival curves below (Fig.10-5), but there is no statistically significant separation of the survival curves for either parameter.

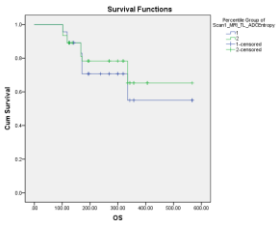
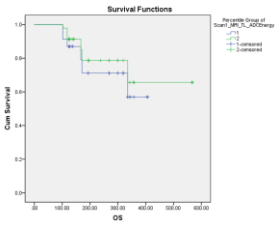
$ADC_{entropy}$	ADC_{energy}
	
Dichotomised at 5.3 Low (≤) – Blue High (>) - Green	Dichotomised at 0.007 Low (≤) – Blue High (>) - Green
p=0.501	p=0.401

Figure 10-5: DW-MRI Per-lesion Analysis; KM OS Analysis; baseline ADC heterogeneity parameters; All patients

Analysing both tumour groups together is perhaps misleading because the prostate cancer patients had a slightly higher baseline ADC_{energy} than the breast cancer patients (p=0.067). KM analysis of the prostate cancer lesions only demonstrates a statistically significant separation of

the OS curves for ADC_{energy}, with a survival advantage noted (p=0.019) for patients with a higher measurement (i.e. more uniform ADC distribution within the VOIs) (Fig.10-6). No association between OS and baseline $ADC_{entropy}$ has been identified.

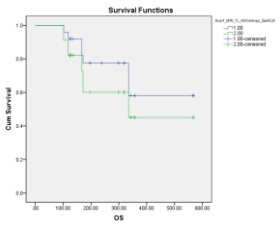
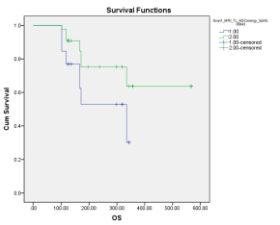
$ADC_{entropy}$ (prostate only)	ADC_{energy} (prostate only)
	
Dichotomised at 5.24 Low (≤) – Blue High (>) - Green	Dichotomised at 0.00643 Low (≤) – Blue High (>) - Green
p=0.214	p=0.019

Figure 10-6: DW-MRI Per-lesion Analysis; KM OS Analysis; baseline ADC heterogeneity parameters; Prostate patients only

10.1.3 %Δ Parameters OS Analysis

10.1.3.1 %ΔADC Parameters

	$\%\Delta ADC_{mean}$	**$\%\Delta ADC_{median}$**
	Cox regression p-value [HR (95%CI)]	
n=87, 25 deaths	**0.24** [1.01 (0.993-1.027)]	**0.308** [1.007 (0.993-1.021)]

Table 10-4: DW-MRI Per-lesion Analysis; %ΔADC Parameters; OS Cox regression analysis

There is no correlation between $\%\Delta ADC_{mean}$ or $\%\Delta ADC_{median}$ and OS (Tbl.10-4).

KM analysis of the data shows statistically significant separation of the $\%\Delta ADC_{mean}$ curves (p=0.037), and for $\%\Delta ADC_{median}$ (p=0.007), showing a survival advantage for lesions with a decrease (or lower increase) in ADC between the scans (Fig.10-7).

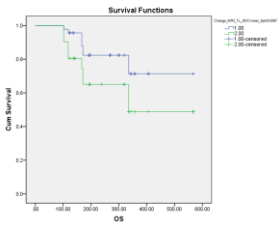
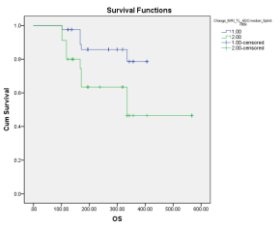
$\%\Delta ADC_{mean}$	$\%\Delta ADC_{median}$
	
Dichotomised at +5.94% Low (≤) – Blue High (>) - Green	Dichotomised at +4.79% Low (≤) – Blue High (>) - Green
p=0.037	p=0.007

Figure 10-7: DW-MRI Per-lesion Analysis; KM OS Analysis; %ΔADC parameters; all lesions

The ADC parameters are difficult to interpret. For the prostate cancer lesions, a lower ADC at baseline was indicative of a longer OS. The breast cancer lesions had on average a higher baseline ADC and at the time of data analysis none of the breast patients had died. Inclusion of the breast lesions for baseline KM analysis showed the clear OS advantage of the breast patients,

and thus it seemed a higher ADC was prognostic. However, by the second scan, a lower ADC was associated with an OS advantage for all the lesions (data not included), suggesting the higher ADC of the breast lesions had decreased. A decrease in ADC would be expected to be associated with an OS benefit, as shown in this analysis.

The breast lesions are contributing to the separation of the KM OS curves. Separation is still demonstrated when analysing just the prostate cancer lesions (Figure 10-8), reaching statistical significance for $\%\Delta ADC_{median}$ (p=0.015), but not quite with $\%\Delta ADC_{mean}$ (p=0.006) $\%\Delta ADC_{median}$ is more discriminating for these lesions.

This analysis concludes utility of the %ΔADC parameters for predicting OS for the prostate cancer lesions.

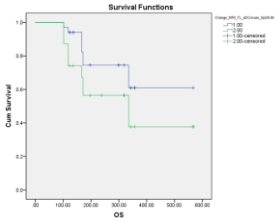
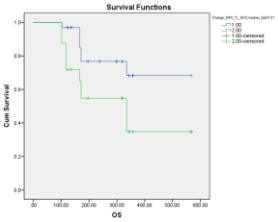
$\%\Delta ADC_{mean}$ (prostate only)	$\%\Delta ADC_{median}$ (prostate only)
	
Dichotomised at +5.94% Low (≤) – Blue High (>) - Green	Dichotomised at +7.51% Low (≤) – Blue High (>) - Green
p=0.06	p=0.015

Figure 10-8: DW-MRI Per-lesion Analysis; KM OS Analysis; %ΔADC parameters; Prostate lesions only

For the prostate cancer patients at least, an increase in ADC confers a poorer OS. It is unclear what change in the bone architecture the ADC is most affected by, but it might be sclerosis is a significant contributing factor; an increase in sclerosis may cause further restriction and a fall in ADC between scans, but may suggest a treatment response and osteoblastic reaction resulting in a prolonged OS. OS analysis of just the breast cancer lesions is not possible as all the patients were alive at the time of data analysis.

10.1.3.2 %ΔtDV

	%ΔtDV
	Cox regression p-value [HR (95%CI)]
n=87, 25 deaths	**0.971** [1 (0.991-1.009)]

Table 10-5: DW-MRI Per-lesion Analysis; %ΔADC Heterogeneity Parameters; OS Cox regression analysis

There is no demonstrable correlation in these patients between the change in size of the lesions (%ΔtDV) and the OS when analysed by Cox regression (Tbl.10-5). Using KM analysis there is no demonstrable association for these lesions between the change in size of the lesions between scans and the OS (Fig.10-9).

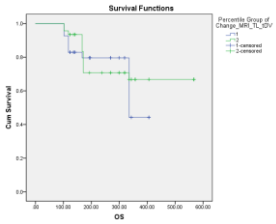
%ΔtDV

Dichotomised at 0 Low (≤) – Blue High (>) - Green
p=0.514

Figure 10-9: DW-MRI Per-lesion Analysis; KM OS Analysis; %ΔtDV

OS analysis of just the prostate cancer lesions also demonstrates no significant separation of the KM survival curves (p=0.817) concluding for these patients the %Δ tDV has no OS predictive utility.

10.1.3.3 %ΔADC Heterogeneity Parameters

	%ΔADC$_{entropy}$	**%ΔADC$_{energy}$**
	Cox regression p-value [HR (95%CI)]	
n=87, 25 deaths	**0.397** [1.031 (0.961-1.105)]	**0.116** [0.993 (0.984-1.002)]

Table 10-6: DW-MRI Per-lesion Analysis; %ΔADC Heterogeneity Parameters; OS Cox regression analysis

There is no demonstrable statistically significant correlation between the %ΔADC heterogeneity parameters when analysed using Cox regression (Tbl.10-6). The KM OS curves can be reviewed below (Fig.10-10):

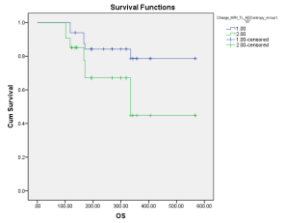
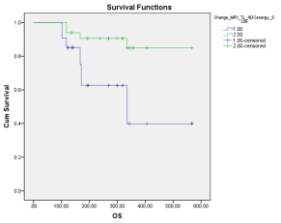
%ΔADC$_{entropy}$	%ΔADC$_{energy}$
	
Dichotomised at -1.127% Low (≤) – Blue High (>) - Green	Dichotomised at +8.239% Low (≤) – Blue High (>) - Green
p=0.023	p=0.002

Figure 10-10: DW-MRI Per-lesion Analysis; KM OS Analysis; %ΔADC Heterogeneity Parameters

Lesions with a more negative change in ADC$_{entropy}$ (≤ -1.127%, p=0.023) and more positive change in ADC$_{energy}$ (> +8.239%, p=0.002) were found in patients with longer OS, i.e. a change in lesions to a more uniform, less heterogeneous state in terms of ADC values. Median OS had not been reached at the time of data analysis.

The separation of the survival curves is maintained when only the prostate cancer lesions are analysed (Figure 10-11), although statistical significance is only reached with %Δ ADC$_{energy}$ (p=0.015).

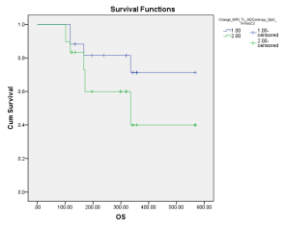
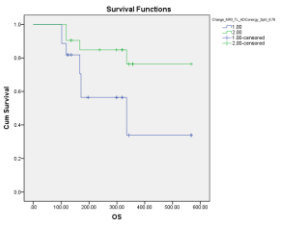
$\%\Delta ADC_{entropy}$ (prostate only)	$\%\Delta ADC_{energy}$ (prostate only)
	
Dichotomised at -2.2% Low (≤) – Blue High (>) - Green	Dichotomised at +8.78% Low (≤) – Blue High (>) - Green
p=0.09	p=0.015

Figure 10-11: DW-MRI Per-lesion Analysis; KM OS Analysis; %ΔADC Heterogeneity Parameters; Prostate patients only

10.2 Lesion Based Analysis – PFS Analysis

10.2.1 Tumour Groups

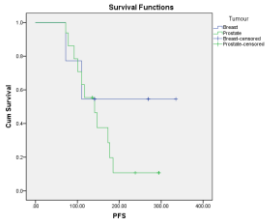
	Tumour Groups
KM plot	
Breast lesions	n/a
Prostate lesions	141 (120.2-161.8)
p-Value (Log Rank)	0.039

Table 10-7: DW-MRI Per-lesion Analysis; PFS Table for tumour groups; All lesions

The breast cancer patients have a PFS advantage over the prostate patients, seen above in the KM analysis (Tbl.10-7, p=0.039). Median PFS for the breast patients has not been reached. Median PFS for the prostate cancer patients is 141 days from initiation of therapy.

10.2.2 Baseline Scan

10.2.2.1 ADC Parameters

	ADC_{mean}	ADC_{median}
	Cox regression p-value [HR (95%CI)]	
n=87, 64 events	**0.295** [0.521 (0.153-1.767)]	**0.519** [0.654 (0.179-2.381)]

Table 10-8: DW-MRI Per-lesion Analysis; Baseline ADC Parameters; PFS Cox regression analysis

There is no demonstrable statistically significant correlation between the magnitude of the baseline ADC parameters measured with these lesions and PFS of the patients when analysed using Cox regression (Tbl.10-8).

KM analysis of the data (Tbl.10-9) shows little separation of the survival plots, but for ADC_{mean} there is an identified 6-days PFS benefit in the lesions with higher ADC_{mean} (p=0.049). There is no statistically significant separation of the ADC_{median} survival curves.

Although this suggests a statistically significant separation of the survival curves, there is only a 6-day difference in PFS. There is no clinically relevant observed difference in PFS identified using baseline ADC parameters.

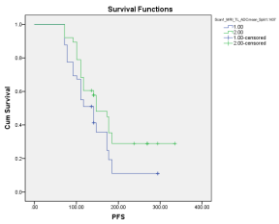
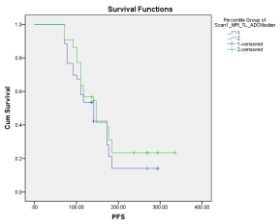
		ADC_{mean}	ADC_{median}
KM plot			
Dichotomisation		1.1437	1.06
Median PFS in days (95% CI)	≤	141 (109.1-172.9)	141 (111.2-170.8)
	>	147 (106-188)	147 (117.7-176.3)
p-Value (Log Rank)		**0.049**	**0.268**

Table 10-9: DW-MRI Per-lesion Analysis; KM PFS Analysis; Baseline ADC_{mean}; All lesions

The baseline imaging showed a higher ADC of the breast cancer lesions, and this had an impact on OS analysis. The breast cancer patients have a PFS advantage. When analysing the prostate lesions alone, there is no significant separation of the KM PFS curves (ADC_{mean} p=0.897, ADC_{median} p=0.524, plots not shown). PFS analysis with only the breast cancer lesions suggests a lower ADC is prognostic of a longer PFS, although this does not reach statistical significance (p=0.106 for both parameters, Tbl.10-10).

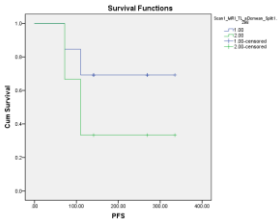
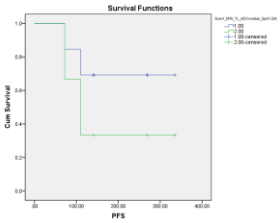
		ADC_{mean} (breast only)	ADC_{median} (breast only)
KM plot			
Dichotomisation		1.298	1.224
Median PFS in days (95% CI)	≤	n/a	n/a
	>	110 (74.9-145.1)	110 (74.9-145.1)
p-Value (Log Rank)		**0.106**	**0.106**

Table 10-10: DW-MRI Per-lesion Analysis; KM PFS Analysis; baseline ADC_{mean}; Only Breast Cancer lesions

The PFS KM analysis including all the lesions is misleading. The breast cancer lesions had a higher baseline ADC than the prostate lesions, and separate analysis suggests a PFS benefit for the breast cancer lesions with a lower baseline ADC. The prostate lesions had a lower baseline ADC, and no difference in PFS has been identified in these patients by KM analysis. However, analysis using baseline ADC suggested an OS benefit for the prostate cancer lesions (see earlier results).

10.2.2.2 tDV

	tDV
	Cox regression p-value [HR (95%CI)]
n=87, 64 events	**0.211** [1.002 (0.999-1.005)]

Table 10-11: DW-MRI Per-lesion Analysis; Baseline tDV; PFS Cox regression analysis

There is no significant correlation between the baseline tDV and the PFS for these lesions (Tbl.10-11). The data has been analysed with the KM PFS method (Tbl.10-12).

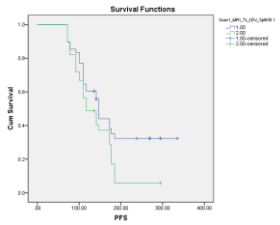
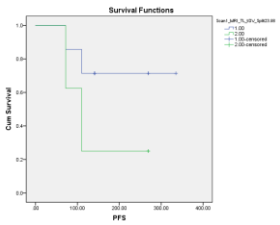
		tDV
KM plot		
Dichotomisation		38.1cm^3
Median PFS in days (95% CI)	≤	147 (139.2-154.8)
	>	117 (87.1-146.9)
p-Value (Log Rank)		**0.031**

Table 10-12: DW-MRI Per-lesion Analysis; KM PFS Analysis; Baseline tDV

There is a significant separation of the survival curves for these lesions. Lesions smaller than 38.1cm^3 have a 30-day PFS benefit (p=0.031) using these data. The breast cancer lesions were smaller than the prostate cancer lesions, and this may simply be identifying the breast cancer lesions (and the breast patients have a PFS advantage).

KM PFS analysis of only the prostate cancer lesions does not demonstrate significant separation of the PFS curves (p=0.177), but smaller breast lesions have a PFS benefit (≤ 23.98cm^3, p=0.038,– see Fig.10-12). This suggests a PFS advantage for smaller breast cancer deposits, which would be biologically anticipated.

tDV (breast lesions only)



Dichotomised at 23.98cm^3
Low (≤) – Blue
High (>) - Green

p=0.038

Figure 10-12: DW-MRI Per-lesion Analysis; KM PFS Analysis; Baseline tDV; Breast lesions

10.2.2.3 Heterogeneity Parameters

	$ADC_{entropy}$	ADC_{energy}
	Cox regression p-value [HR (95%CI)]	
n=87, 64 events	**0.561** [1.339 (0.501-3.578)]	**0.201** [0 (0-643663.969)]

Table 10-13: DW-MRI Per-lesion Analysis; Baseline ADC Heterogeneity Parameters; PFS Cox regression analysis

There is no significant correlation in the data for these lesions between baseline ADC heterogeneity parameters and PFS using Cox regression (Tbl.10-13). The data has been dichotomised for KM analysis (Tbl.10-14).

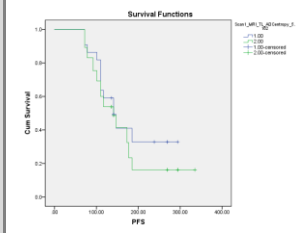
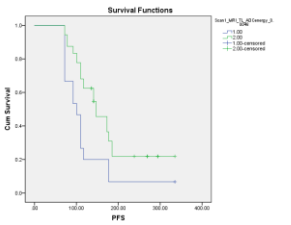
		$ADC_{entropy}$	ADC_{energy}
KM plot			
Dichotomisation		5.162	0.0048
Median PFS in days (95% CI)	≤	141 (105.8-176.2)	101 (77-125)
	>	141 (116.5-165.5)	147 (120.8-173.2)
p-Value (Log Rank)		0.340	0.004

Table 10-14: DW-MRI Per-lesion Analysis; KM PFS Analysis; Baseline ADC heterogeneity parameters

There is no significant separation of the survival curves for $ADC_{entropy}$ for these data. ADC_{energy} however shows a statistically significant separation, with a PFS benefit of median 46 days in those patients with baseline ADC_{energy} of greater than 0.0048 (p=0.004) suggesting the more uniform lesions (higher ADC_{energy}) progressed later than more heterogeneous lesions.

The prostate lesions had a trend towards having a higher ADC energy on baseline imaging. The prostate cancer patients had a shorter PFS, and the benefit shown in the KM PFS analysis above is most from the breast cancer lesions. Indeed, PFS analysis of only the prostate cancer lesions shown no significant separation of the PFS curves (dichotomisation 0.0066, p=0.710).

The same analysis for the breast cancer lesions is shown in Figure 10-13, demonstrating a significant separation of the KM PFS curves, with a clear benefit for lesions with a higher baseline ADC_{energy} (>0.0049).

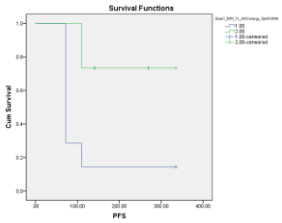
ADC_{energy} (breast lesions only)

Dichotomised at 0.0049 Low (≤) – Blue High (>) - Green
p=0.001

Figure 10-13: DW-MRI Per-lesion Analysis; KM PFS Analysis; Baseline ADC_{energy}; Breast lesions only

10.2.3 %Δ Parameters

10.2.3.1 %ΔADC Parameters

	$\%\Delta ADC_{mean}$	**$\%\Delta ADC_{median}$**
	Cox regression p-value [HR (95%CI)]	
n=87, 64 events	**0.903** [0.999 (0.988-1.01)]	**0.781** [0.999 (0.989-1.008)]

Table 10-15: DW-MRI Per-lesion Analysis; %ΔADC Parameters; PFS Cox regression analysis

Cox regression analysis of the %ΔADC parameters has not identified significant correlation between the size of change and the length of PFS (Table 10-15).

KM PFS analysis has not identified significant separation of the survival curves (Table 10-16).

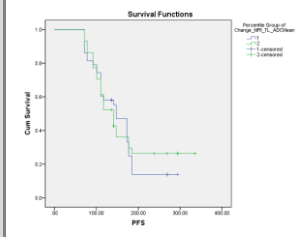
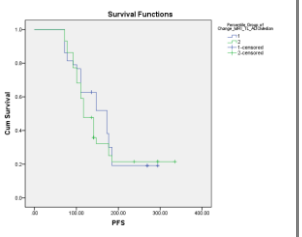
		$\%\Delta ADC_{mean}$	**$\%\Delta ADC_{median}$**
KM plot			
Dichotomisation		+5.71%	+5.84%
Median PFS in days (95% CI)	**≤**	147 (124.7-169.3)	173 (133.8-212.2)
	>	141 (116.5-165.5)	117 (97.4-136.6)
p-Value (Log Rank)		**0.762**	**0.395**

Table 10-16: DW-MRI Per-lesion Analysis; KM PFS Analysis; %ΔADC parameters; All lesions

The PFS KM analysis for the separate tumour groups can be reviewed in Tables 10-17 and 10-18.

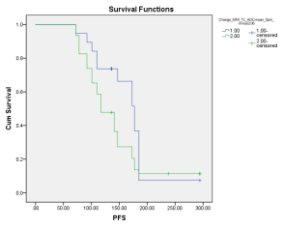
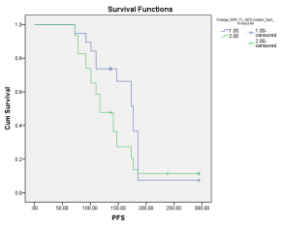
		%ΔADC$_{mean}$ (prostate only)	%ΔADC$_{median}$ (prostate only)
KM plot			
Dichotomisation		-2.06	-3.49
Median PFS in days (95% CI)	≤	177 (151.7-202.3)	177 (151.7-202.3)
	>	117 (96.9-137.1)	117 (96.9-137.1)
p-Value (Log Rank)		**0.088**	**0.088**

Table 10-17: DW-MRI Per-lesion Analysis; KM PFS Analysis; %ΔADC parameters; Prostate lesions only

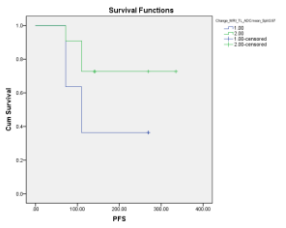
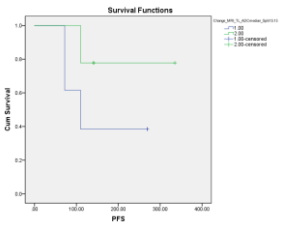
		%ΔADC$_{mean}$ (breast only)	%ΔADC$_{median}$ (breast only)
KM plot			
Dichotomisation		+3.97%	+13.13%
Median PFS in days (95% CI)	≤	110 (70.4-149.6)	110 (66.5-153.5)
	>	n/a	n/a
p-Value (Log Rank)		**0.078**	**0.052**

Table 10-18: DW-MRI Per-lesion Analysis; KM PFS Analysis; %ΔADC parameters; Prostate lesions only

Neither of the parameters reaches statistical significance for either tumour group, but there is certainly a trend. The tumour groups seem to have different characteristics. A fall in ADC between scans is associated with a PFS benefit for the prostate cancer lesions, but an increase in ADC is associated with a PFS benefit for the breast cancer lesions. This reflects a difference in the pathophysiology of a treatment response on the metastases and surrounding bone between the tumour types.

10.2.3.2 %ΔtDV

	%ΔtDV
	Cox regression p-value [HR (95%CI)]
n=87, 64 events	**0.478** [0.998 (0.993-1.003)]

Table 10-19: DW-MRI Per-lesion Analysis; %ΔtDV; PFS Cox regression analysis

There is no demonstrable correlation between the size of change in tDV and the PFS for these lesions (Table 10-19). KM analysis can be reviewed in the table below (Table 10-20).

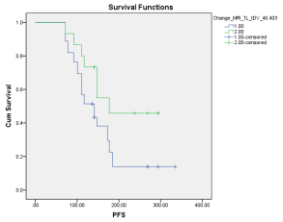
		%ΔtDV
KM plot		
Dichotomisation		+46.403%
Median PFS in days (95% CI)	≤	141 (114.3-167.7)
	>	177 (n/a)
p-Value (Log Rank)		**0.027**

Table 10-20: DW-MRI Per-lesion Analysis; KM PFS Analysis; %ΔtDV; All lesions

Lesions with a greater than +46% increase in volume are associated with a 36-day PFS benefit (p=0.027) when all the lesions are considered together. Tumour-specific analysis can be reviewed in Tables 10-21 and 10-22.

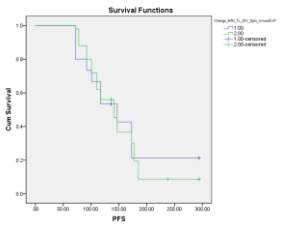
		%ΔtDV (prostate lesions only)
KM plot		
Dichotomisation		-20.47%
Median PFS in days (95% CI)	≤	147 (94.2-199.8)
	>	141 (119.4-162.6)
p-Value (Log Rank)		**0.819**

Table 10-21: DW-MRI Per-lesion Analysis; KM PFS Analysis; %ΔtDV; Breast lesions only

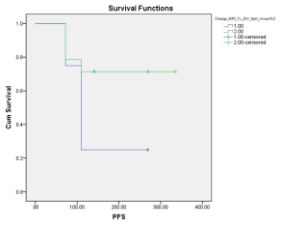
		%ΔtDV (breast lesions only)
KM plot		
Dichotomisation		-19.2%
Median PFS in days (95% CI)	≤	110 (87.2-132.8)
	>	n/a
p-Value (Log Rank)		**0.071**

Table 10-22: DW-MRI Per-lesion Analysis; KM PFS Analysis; %ΔtDV; Breast lesions only

There is no association between the change in tDV and PFS for the prostate cancer lesions (p=0.819). A PFS benefit is seen for the breast cancer lesions with an increase in tDV. This is an unexpected finding, and the separation of the PFS curves does not reach statistical significance. There is a different pathophysiology demonstrated in this per-lesion analysis between the tumour groups, and this finding for the breast cancer lesions most likely reflects a change in the bone metastasis or the surrounding bone. The breast cancer patients in this study had a PFS benefit, and significant OS benefit, and therefore the measured increase in tDV is unlikely to represent an increase in the size of the tumour deposit, and tDV may not be a true measure of just the tumour deposit in the bone, and is more likely to represent the tumour and the surrounding bony architecture.

10.2.3.3 %ΔADC Heterogeneity Parameters

	$\%\Delta ADC_{entropy}$	$\%\Delta ADC_{energy}$
	Cox regression p-value [HR (95%CI)]	
n=87, 64 events	**0.841** [1.004 (0.966-1.044)]	**0.201** [0.999 (0.996-1.001)]

Table 10-23: DW-MRI Per-lesion Analysis; %ΔADC Heterogeneity Parameters; PFS Cox regression analysis

There is no significant correlation for these lesions between the change in the ADC heterogeneity parameters and the length of PFS (Tbl.10-23). The data has been dichotomised for KM PFS analysis (Tbl.10-24).

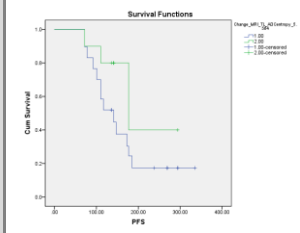
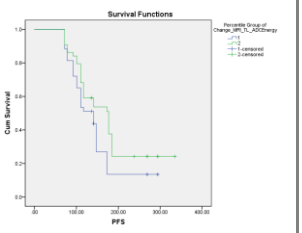
		$\%\Delta ADC_{entropy}$	$\%\Delta ADC_{energy}$
KM plot			
Dichotomisation		+5.564%	-0.03%
Median PFS in days (95% CI)	≤	141 (115.5-166.5)	141 (112.9-169.1)
	>	177 (81.8-272.2)	177 (135-219)
p-Value (Log Rank)		**0.074**	**0.053**

Table 10-24: DW-MRI Per-lesion Analysis; KM PFS Analysis; %ΔADC Heterogeneity Parameters; All lesions

Lesions with a more negative change in $ADC_{entropy}$ between scans have a better PFS, although this does not quite reach statistical significance (p=0.074). There is a 36-day PFS benefit in lesions with a greater positive change in ADC_{energy} (p=0.053). These data therefore suggest a change towards homogeneity confers a PFS benefit.

The tumour subgroup analyses show contradictory findings depending on the tumour type being assessed (Tbls.10-25/10-26). For the prostate cancer lesions, there is significant separation of both PFS curves ($\%\Delta ADC_{entropy}$ p=0.006, $\%\Delta ADC_{energy}$ p=0.0002) showing at least a 60-day PFS benefit for less heterogeneous lesions (decrease in $ADC_{entropy}$ and increase in ADC_{energy}). The exact opposite is demonstrated by the breast cancer lesions, where an increase in $ADC_{entropy}$ and decrease in ADC_{energy} is associated with a PFS benefit (p=0.052).

The number of breast patients is small, and the log rank analysis does not quite reach the level of statistical significance, particularly when repeated testing is considered, but this suggests a fundamental difference at the physiological level about the behaviour of the prostate and breast cancer lesions in these patients.

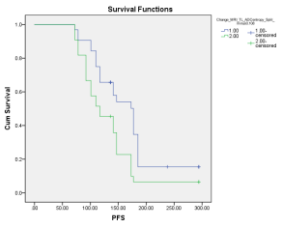
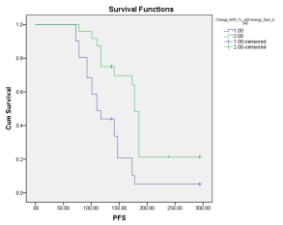
		$\%\Delta ADC_{entropy}$ (prostate lesions only)	$\%\Delta ADC_{energy}$ (prostate lesions only)
KM plot			
Dichotomisation		-0.108%	+6.245%
Median PFS in days (95% CI)	≤	177 (149.3-204.7)	110 (93.3-126.7)
	>	117 (85.9-148.1)	177 (170.9-183.1)
p-Value (Log Rank)		**0.006**	**0.0002**

Table 10-25: DW-MRI Per-lesion Analysis; KM PFS Analysis; %ΔADC Heterogeneity Parameters; Prostate lesions only

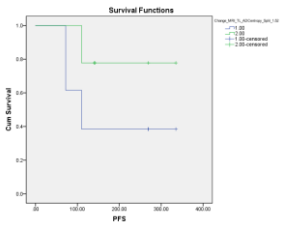
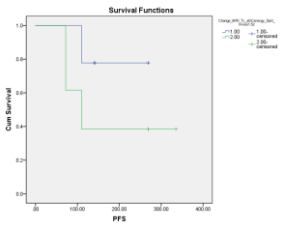
		$\%\Delta ADC_{entropy}$ (breast lesions only)	$\%\Delta ADC_{energy}$ (breast lesions only)
KM plot			
Dichotomisation		+1.52%	-1.52%
Median PFS in days (95% CI)	≤	110 (66.5-153.5)	n/a
	>	n/a	110 (66.5-153.5)
p-Value (Log Rank)		**0.052**	**0.052**

Table 10-26: DW-MRI Per-lesion Analysis; KM PFS Analysis; %ΔADC Heterogeneity Parameters; Breast lesions only

10.3 Discussion

		ADC_{mean}	ADC_{median}	tDV	$ADC_{entropy}$	ADC_{energy}
Baseline Scan	**↑OS**	PCa: ↓ p=0.047 BCa: n/a	PCa: ↓ p=0.005 BCa: n/a	X	X	PCa: ↑ p=0.019
	↑PFS	PCa: X BCa: ?↓ p=0.106	PCa: X BCa: ?↓ p=0.106	X	X	PCa: X BCa: ↑ p=0.001
%Δ	**↑OS**	PCa:↓ p=0.06 BCa: n/a	PCa:↓ p=0.005 BCa: n/a	X	↓ p=0.023	↑ p=0.002
	↑PFS	PCa: ↓ p=0.088 BCa:↑ p=0.078	PCa: ↓ p=0.088 BCa:↑ p=0.052	PCa: X BCa:↑ p=0.07	PCa: ↓ p=0.006 BCa: ↑ p=0.052	PCa: ↑ p=0.0002 BCa: ↓ p=0.052

Table 10-27: DW-MRI Per-lesion Analysis; All Survival Analyses; Summary Table

Useful prognostic information has been identified by imaging tumour deposits with DW-MRI. For prostate cancer, a negative correlation between ADC and Gleason grading has been identified [220]. Histological correlates of a poor response have been identified for other tumours. Breast primary tumours with a low ADC at baseline are associated with a higher tumour grade and invasiveness, and with a worse prognosis[236]. A high baseline ADC in head and neck primary tumours predicts a poorer outcome[217], and is associated histologically with poor prognostic features (higher stromal content, low cellularity (suggesting lower proliferation), and micronecrosis), and poorly differentiated tumours have a lower ADC than well-differentiated ones[237-239].

The per-lesion response analysis for this study was not able to demonstrate predictive benefit using the baseline ADC parameters, but the %Δ ADC parameters suggested an increase in ADC with treatment predicted a clinical treatment response for the prostate cancer lesions.

None of the breast cancer patients had died at the time of data analysis for this study, and this OS advantage (p=0.002) and PFS advantage (p=0.039), for the breast cancer patients has impacted interpretation of the data.

Lower baseline ADC parameters (mean and median) predict OS (p=0.047 and 0.005 respectively, but median OS was not reached for comparison) for the prostate lesions. Other studies have shown a lower baseline ADC is predictive of a treatment response, and OS might similarly be expected to be prolonged. There is little in the literature to support this supposition. Sclerotic bone metastases have a better prognosis[235], but sclerosis is not the confirmed cause of the lower ADC in these lesions. Sclerosis does return a lower ADC measurement [234]. ADC_{median} is more predictive for these lesions.

The baseline ADC parameters have not shown a statistically significant PFS prognostic benefit.

This difference between the tumour groups persists through to analysis of the %ΔADC parameters, where a fall in ADC between scans is associated with a non-statistically significant PFS benefit for prostate cancer patients ($ADC_{mean/median}$ p=0.088 for both), but an increase in ADC

between scans of the breast cancer lesions confers a non-statistically significant PFS benefit ($ADC_{mean/median}$ p=0.078/0.052). There is currently no histological correlate of what the changes in ADC are measuring in these tumour metastases, but there are fundamental differences in the pathophysiology of the bone metastases from breast and prostate cancer. Inclusion of mixed bone pathologies in future studies may mask or distort the findings; bone metastases differ in phenotypic and biological behaviour.

There has been some evidence to suggest the volume of bony metastatic disease has prognostic benefit. Perez-Lopez et al. reported in 2016 a study of 43 metastatic prostate cancer patients, showing the baseline tDV^{WB} was associated with OS, and a greater tumour burden in the bone had a negative impact on OS[240]. Vargas et al. demonstrated an increased number of prostate cancer bone metastases was associated with a shorter survival[241] but it is not clear if this is directly analogous to tDV. The bone scan index is a measure of the proportion of the skeleton replaced by metastatic disease, and has been shown to predict prostate cancer survival, with a poorer survival from a higher burden[222,223,242].

No PFS prognostic utility has been identified using tDV for the prostate cancer lesions. The breast cancer lesions tell a different tale: smaller (lower tDV) breast cancer metastases are associated with an improved PFS (scan1/scan2 p=0.038/0.0002). The %ΔtDV for the breast cancer lesions does not quite reach statistical significance, but there is a trend towards a PFS benefit for lesions with an increase in tDV between the scans (p=0.071). This is clearly an unexpected finding, and an explanation is difficult. This could be a statistical phenomenon, suggesting there may be a difference, or simply random error. The tDV may not just measuring viable tumour volume, but might also be measuring pathophysiological changes in the bony architecture; the responses and changes with treatment may be causing a halo-effect of change within the surrounding bone, volume on the high b-value image and resulting in an increased VOI volume. This again indicates the impact of the underlying bone metastasis pathophysiology on the quantification parameters.

It has already been described how the ADC heterogeneity parameters have been associated with markers of poorer prognosis for several tumours, including prostate cancer[176] and breast cancer[172]. A higher entropy and lower energy (i.e. a more heterogeneous ADC voxel distribution within the VOIs) is associated with more poor prognosis histological features, and associations with OS and PFS it would be expected.

ADC_{energy} seems to have more prediction utility for these patients. A higher baseline ADC_{energy} predicts better OS for the prostate cancer patients (p=0.019). Both %Δ ADC heterogeneity parameters have successfully predicted the OS for the prostate cancer lesions: a decrease in $ADC_{entropy}$ and increase in ADC_{energy} between scans shows an OS benefit for the prostate cancer lesions (p=0.023 and 0.002 respectively). A change towards a more homogeneous ADC voxel distribution confers an OS benefit. The underlying bone physiology being measured is unknown, but this may represent a return to a more homogeneous normal bone architecture, or a move towards sclerosis, which may result in a more homogeneous ADC voxel distribution. There is no histological correlate in the literature.

The ADC heterogeneity parameters also show performance for PFS analysis for these patients. ADC_{energy} offers more prognostic value. For the breast cancer lesions, only a higher ADC_{energy} at baseline is associated with a PFS benefit (p=0.001). No such benefit was identified for the prostate cancer lesions. The PFS of the %Δ in ADC parameters is more difficult to interpret as the underlying tumour type demonstrates conflicting outcomes. For the breast cancer lesions, an increase in $ADC_{entropy}$ and decrease in ADC_{energy} (i.e. towards heterogeneity) is associated with a PFS benefit (p=0.006 and 0.0002, respectively). However, for the prostate cancer lesions, a decrease in $ADC_{entropy}$ and increase in ADC_{energy} (i.e. towards homogeneity) is associated with a PFS benefit (p=0.052 for both). The changes in the parameters for the prostate cancer lesions seem to fit with expectation, and agree with the OS analysis, where survival is indicated by a move of the ADC distribution within VOIs towards homogeneity. The move towards heterogeneity for PFS benefit of the breast cancer lesions suggests a different underlying pathophysiological response of the tumour and bony architecture to the endocrine therapy.

In summary, prognostication of OS has been demonstrated with a lower baseline ADC and higher ADC_{energy} (showing more homogeneity of ADC voxels) of the prostate cancer lesion. Breast and prostate metastases show different parameter characteristic changes with treatment; for the prostate cancer lesions a fall in ADC is associated with a PFS benefit, but the opposite is demonstrated for the breast cancer lesions. Similar conflicting data is shown for PFS prognostication using the %Δ heterogeneity parameters. This suggests the different biology of the tumour types is affecting quantification different.

Chapter 11 DW-MRI Per-Patient Analysis – Response Analysis

The methods of this chapter and the statistical approaches for this section can be reviewed in Chapter 2.3.10.

11.1 Baseline Scan

11.1.1 Tumour Group

n=20 (BCa-5, PCa-15). 18 patients had data for response analysis (BCa-5 (PD-2, Response 3/5(60%), PCa-13 (PD-11, Response 2/13(15.4%)).

The breast cancer patients had a higher response rate.

11.1.2 ADC Parameters

There is a clear demonstrable difference in the distributions (Tbl.11-1) between the two parameters (paired samples t-test: when t(19)=7.316, p=<0.00001). Although ADC_{mean} and ADC_{median} are correlated, there is a significant difference in their distributions. The distributions are wider for the prostate cancer patients, suggesting more intra-patient heterogeneity. The distribution means of the ADC parameters for the prostate cancer metastases are significantly lower than those for breast cancer (Mann Whitney U tests: $ADC_{mean,}$ p=0.028; ADC_{median} p=0.058 suggesting a trend).

The box-plots suggests a lower baseline ADC parameters might predict PD, but this difference is not statistically significant (independent samples t-test: ADC_{mean}, where t(16)=-1.653, *p*=0.118; ADC_{median}, p=0.154).

The breast cancer patients have a higher baseline ADC, and these patients have a better treatment response rate.

Tbl.11-2 summarises the tumour-specific response analysis, comparing the groups with independent sample Mann Whitney U tests (data not recorded here for brevity). This analysis is complicated by the small number of patients. Only 2 prostate cancer patients had a treatment response; 2 of the 5 breast cancer patients had PD. The box-plots demonstrate no statistically significant difference between the response groups, but a treatment response is suggested by a higher baseline ADC. The per-lesion analysis suggested an association between a higher baseline ADC and a treatment response, but did not reach statistical significance.

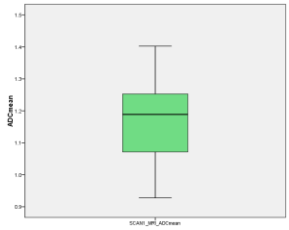
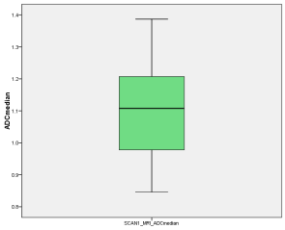
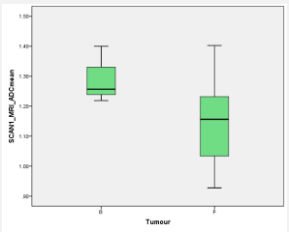
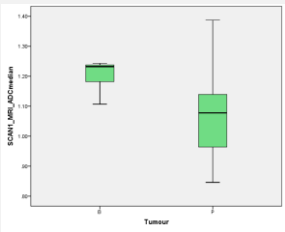
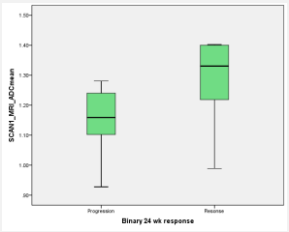
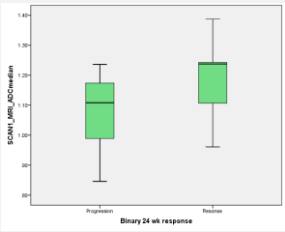
	ADC_{mean}	**ADC_{median}**
	Mean ($x10^{-3}$ mm^2/s) (min-max) [median; SD]	
	1.17 (0.93-1.4) [1.19; 0.14]	**1.1** (0.85-1.39) [1.11; 0.14]
		
	Breast **n=5**	**Prostate** **n=15**	**Breast** **n=5**	**Prostate** **n=15**
	1.29 (1.22-**1.**4) [1.26; 0.08]	**1.14** (0.93-1.4) [1.16; 0.13]	**1.2** (1.11-1.24) [1.23; 0.06]	**1.06** (0.85-1.39) [1.08; 0.15]
Tumour				
	t(18)=2.398, p=0.028		t(18)=2.024, p=0.058	
	PD **n=13**	**Response** **n=5**	**PD** **n=13**	**Response** **n=5**
	1.16 (0.93-1.28) [1.16; 0.11]	**1.27** (0.99-1.4) [1.33; 0.17]	**1.08** (0.85-1.24) [1.11; 0.12]	**1.19** (0.96-1.39) [1.24; 0.16]
Response				
	p=0.118		p=0.154	

Table 11-1: DW-MRI Per-Patient Analysis; Baseline Scan ADC Parameters; Descriptive statistics and box-plots; All Patients (ADC expressed as 10^{-3} mm^2/s); Tumour groups (B=Breast cancer, P=Prostate Cancer)

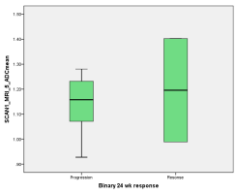
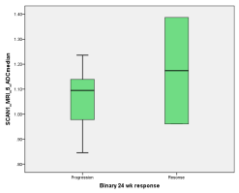
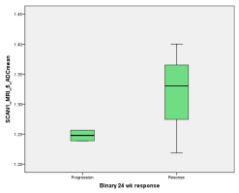
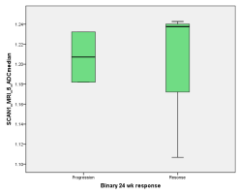
Prostate		**Breast**	
ADC_{mean}	ADC_{median}	ADC_{mean}	ADC_{median}
			
p=1.0	p=0.769	p=1.0	p=1.0

Table 11-2: DW-MRI Per-Patient Analysis; Response Analysis; Baseline ADC Parameters; Tumour Subgroup Analysis

11.1.3 Volumetric tDV

n=20 (5 Breast, 15 Prostate)

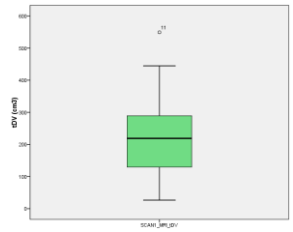
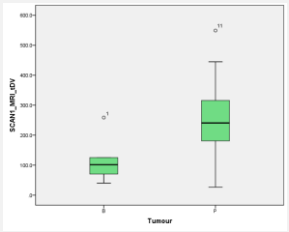
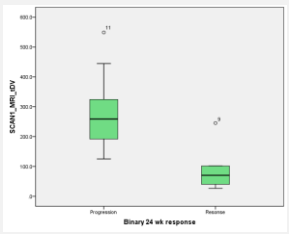
	tDV (cm^3)
	Mean (cm^3) (min-max) [median; SD]
	223.86 (26.72-549.05) [218.82; 133.18]
	
	Breast **n=5** / **Prostate** **n=15**
	119.28 (40.16-258.82) [101.75; 84.35] / **258.72** (26.72-549.05) [240.82; 129.73]
Tumour Groups	 p=0.033
	PD **n=13** / **Response** **n=5**
	278.88 (125.24-549.05) [258.82; 121.74] / **96.91** (26.72-245.47) [70.45; 87.95]
Response Group	 p=0.007

Table 11-3: DW-MRI Per-Patient Analysis; Baseline Scan tDV; Descriptive statistics and box-plots; All Patients; Tumour groups (B=Breast cancer, P=Prostate Cancer)

The prostate cancer patients have larger tumour volumes selected for quantification than the breast cancer patients (Tbl.11-3) (Mann-Whitney U test: p=0.033). Patients with a larger tDV were more likely to have PD by 24-weeks (Mann-Whitney U test: p=0.007).

The prostate patients had larger volume disease, and were more likely to have PD. The subgroup analyses are plotted in Table 11-4, with independent samples Mann Whitney U tests used to compare the response groups (raw data not included for brevity).

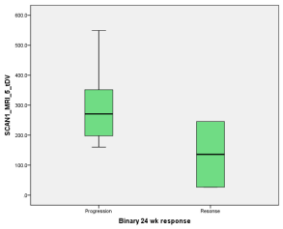
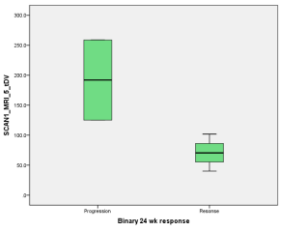
tDV	
Prostate	Breast
	
p=0.308	p=0.2

Table 11-4: DW-MRI Per-Patient Analysis; Response Analysis; Baseline tDV; Tumour Subgroup Analyses

Statistical analysis of the tumour subgroup response groups is difficult because of the small number of patients as previously described. None of these analyses meet statistical significance, but there seems to be a trend towards patients with larger tDV measurements being more likely to have PD.

The per-lesion analysis demonstrated no predictive utility of the baseline tDV, only suggesting tDV identified the tumour group and therefore the inherent treatment response superiority of the breast cancer patients included in this study.

11.1.4 ADC Heterogeneity Parameters

Table 11-5 summaries the data for the ADC heterogeneity parameters. There is no demonstrably significant statistical difference in the distribution of $ADC_{entropy}$ between breast cancer and prostate cancer bone metastases (Mann-Whitney U test: p=0.395).

The population distribution mean of $ADC_{entropy}$ is lower in patients with a treatment response (see Table 11-5), suggesting the lesions with a more uniform ADC distribution are the lesions with a resposne. This difference is not statistically significant (independent samples t-test: when t(16)=0.863, p=0.562)

Patients with a treatment response had bone metastases seem to have a slightly higher ADC_{energy} distribution (i.e. less heterogeneous tumours). This small difference is not statistically significant (Mann-Whitney U test: p=0.775).

Particularly for $ADC_{entropy}$ there is the suggestion of a difference in the parameter between the response groups, with more homogeneous lesions being more likely to have a treatment response. This is not proven statistically; there may be no real difference or the sample size was too small to statistically detect the difference.

Table 11-6 displays the tumour-specific response analysis, with Mann-Whitney U tests for comparisons. None of the statistical comparisons of the tumour subgroup response analyses reach statistical significance, but there are only a small number of patients. The patients with a treatment response seem to have a lower baseline $ADC_{entropy}$ and higher ADC_{energy}, i.e. are more homogeneous before treatment. The per-lesion analysis also demonstrated no difference in the baseline ADC heterogeneity parameters between the response groups.

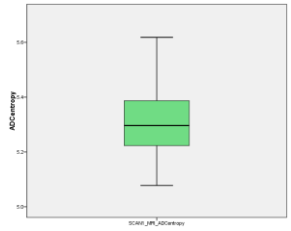
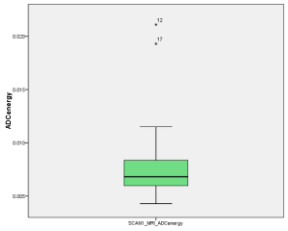
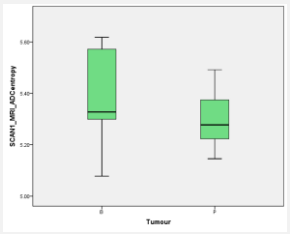
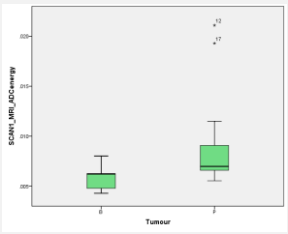
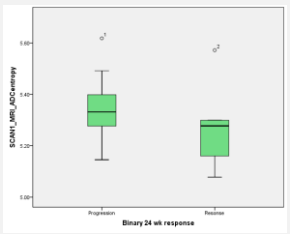
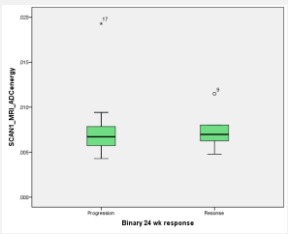
	$ADC_{entropy}$	ADC_{energy}
	Mean (min-max) [median; SD]	
	5.32 (5.08-5.62) [5.3; 0.14]	**0.0083** (0.0043-0.0211) [0.0068; 0.0044]
		
	Breast n=5	**Prostate n=15**	**Breast n=5**	**Prostate n=15**
	5.38 (5.08-5.62) [5.33; 0.22]	**5.3** (5.15-5.49) [5.28; 0.11]	**0.0059** (0.004-0.01) [0.0062; 0.0015]	**0.0091** (0.0055-0.0211) [0.007; 0.0048]
Tumour Groups				
	p=0.395		p=0.06	
	PD n=13	**Response n=5**	**PD n=13**	**Response n=5**
	5.34 (5.15-5.62) [5.33; 0.13]	**5.28** (5.08-5.57) [5.28; 0.19]	**0.0077** (0.0043-0.0193) [0.0067; 0.0037]	**0.0075** (0.0048-0.0115) [0.007; 0.0025]
Response				
	p=0.562		p=0.467	

Table 11-5: DW-MRI Per-Patient Analysis; Baseline Scan ADC Heterogeneity Parameters; Descriptive statistics and box-plots; All Patients; Tumour groups (B=Breast cancer, P=Prostate Cancer)

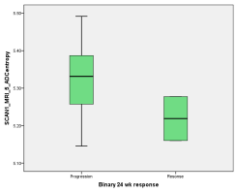
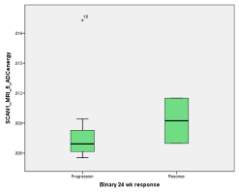
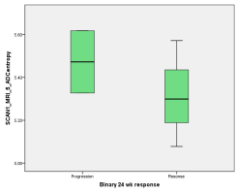
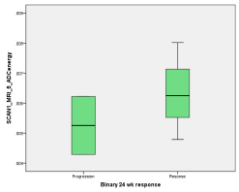
Prostate		**Breast**	
$ADC_{entropy}$	ADC_{energy}	$ADC_{entropy}$	ADC_{energy}
			
p=0.308	p=0.410	p=1.0	p=0.4

Table 11-6: DW-MRI Per-Patient Analysis; Response Analysis; Baseline ADC Heterogeneity Parameters; Tumour Subgroup Analysis

11.1.5 Baseline Parameters Correlations

Tbl.11-7 summarises pairwise correlation analysis between the baseline scan parameters (analysed using Spearman Rank correlation tests).

	ADC_{mean}	**ADC_{median}**	**tDV**	**$ADC_{entropy}$**	**ADC_{energy}**
ADC_{mean}	1	-	-	-	-
ADC_{median}	0.95 (0)	1	-	-	-
tDV	-0.35 (0.125)	-0.29 (0.215)	1	-	-
$ADC_{entropy}$	0.55 (0.0112	0.56 (0.01)	0.21 (0.37)	1	-
ADC_{energy}	-0.47 (0.036)	-0.49 (0.03)	0.09 (0.714)	-0.52 (0.018)	1

Table 11-7: DW-MRI Per-lesion Analysis; Baseline Parameters; Correlations (and p-values) using Spearman Rank analysis; All lesions

A strong correlation is identified between ADC_{mean} and ADC_{median} parameters, as would be anticipated as they are describing the average of the same VOIs. This was also identified with the per-lesion analysis. There is no correlation between tDV and the other parameters, just as identified with the per-lesion analysis. A moderate negative correlation exists between ADC_{energy} and the ADC parameters. The ADC heterogeneity parameters have a close negative correlation. These correlations are similar to those identified with the per-lesion analysis.

11.2 Second Scan

n= 19 (5 Breast, 14 Prostate)

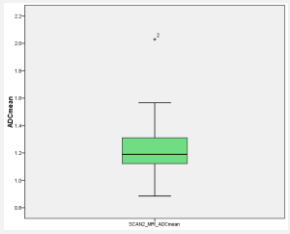
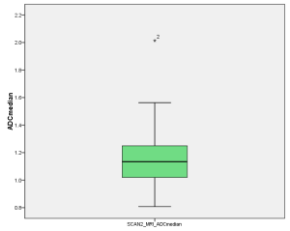
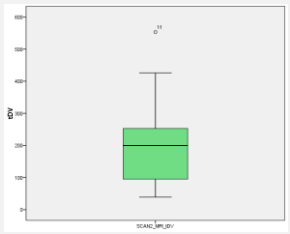
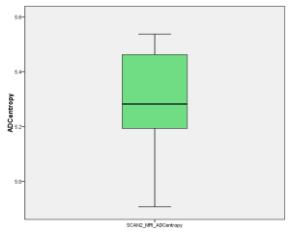
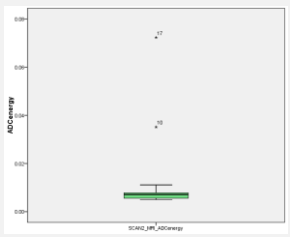
	VOI **Mean ($x10^{-3}$ mm^2/s)** (min-max) [median; SD]	
ADC_{mean} **10^{-3} mm^2/s**	**1.24** (0.89-2.03) [1.19; 0.25]	
ADC_{median} **10^{-3} mm^2/s**	**1.17** (0.81-2.01) [1.13; 0.28]	
tDV **cm^3**	**209.59** (39.13-553.06) [199.71; 143.45]	
$ADC_{entropy}$	**5.29** (4.91-5.54) [5.28; 0.19]	
ADC_{energy}	**0.0118** (0.005-0.0723) [0.0071; 0.0161]	

Table 11-8: DW-MRI Per-Patient Analysis; Second Scan Parameters; Descriptive statistics and box-plots

11.3 Percentage Change (%Δ) Between Scans

n= 19 (5 Breast, 14 Prostate) (18 with data for response analysis, 5 breast, 13 prostate)

11.3.1 %Δ ADC Parameters

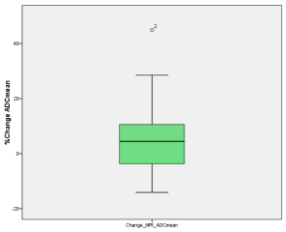
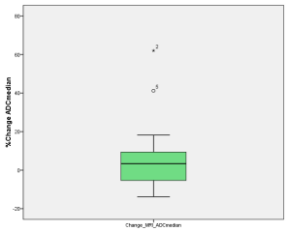
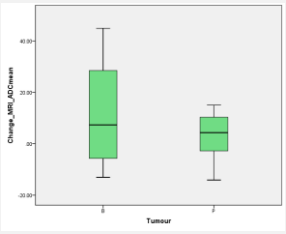
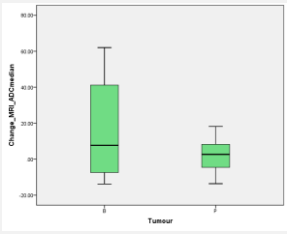
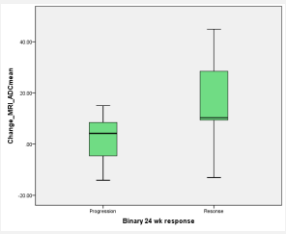
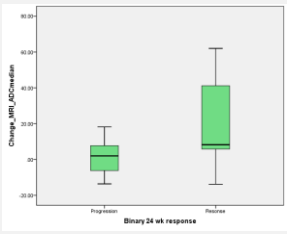
	%ΔADC$_{mean}$		%ΔADC$_{median}$	
	Mean % Change (min-max) [median; SD]			
	5.65 (-14.11-44.92) [4.45; 13.97]		**6.45** (-13.82-62.02) [3.36; 18.52]	
				
	Breast **n=5**	**Prostate** **n=14**	**Breast** **n=5**	**Prostate** **n=14**
	12.4 (-13.04-44.92) [7.33; 24.06]	**3.24** (-14.11-15.03) [4.32; 8.25]	**17.94** (-13.82-62.02) [7.69; 32.56]	**2.35** (-13.6-18.19) [2.68; 8.95]
Tumour				
	t(4.341)=0.834, p=0.448		t(4.218)=1.056, p=0.348	
	PD **n=13**	**Response** **n=5**	**PD** **n=13**	**Response** **n=5**
	2.31 (-14.11-15.03) [4.2; 8.4]	16.02 (-13.04-44.92) [10.32; 21.86]	**1.75** (-13.6-18.19) [2; 9.48]	**20.69** (-13.82-62.02) [8.23; 30.39]
Response				
	p=0.018		p=0.001	

Table 11-9: DW-MRI Per-Patient Analysis; %Δ ADC Parameters; Descriptive statistics and box-plots; All Patients; Tumour groups (B=Breast cancer, P=Prostate Cancer)

Table 11-9 summaries the data for the %Δ ADC parameters. Independent samples t-tests were used to compare the %ΔADC$_{mean}$ distributions for the breast and prostate cancer patients, but no significant difference has been identified (see Tbl.11-9 for results).

The patients with a subsequent treatment response had a more positive change in the ADC parameters between scans. This is what would be anticipated with a treatment response, and

this difference is statistically significant in this analysis (%ΔADC$_{mean}$, when t(16)=-1.985, p=0.018; %ΔADC$_{median,}$ where t(16)=-2.084, p=0.001).

Tumour subgroup analysis is complicated for these data due to the small number of prostate cancer patients with a treatment response (n=2), and the small number of breast cancer patients. The raw data is not displayed, but the plots in Tbl.11-10 display the tumour subgroup response analysis, using Mann-Whitney U tests to compare the distributions:

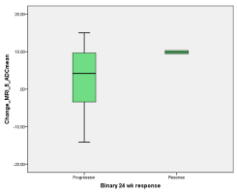
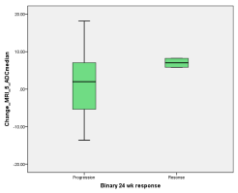
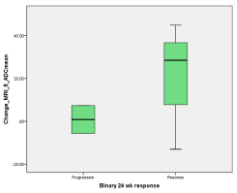
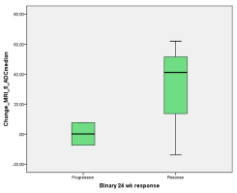
Prostate		Breast	
%ΔADC$_{mean}$	%ΔADC$_{median}$	%ΔADC$_{mean}$	%ΔADC$_{median}$
			
p=0.410	p=0.410	p=1.0	p=1.0

Table 11-10: DW-MRI Per-Patient Analysis; Response Analysis; %Δ ADC Parameters; Tumour Subgroup Analysis

As anticipated with such small numbers none of the differences in the distributions reach statistical significance. The largest influence on the response analysis is the influence of the tumour type. While not statistically significant, the breast cancer patients in this study population had, on average, a larger increase in ADC between scans, and were more likely to have a treatment response.

The per-lesion analysis suggested a treatment response was associated with an increase in ADC between the scans (particularly %Δ ADC$_{mean}$), but this was only observed for prostate cancer bone metastases. There are too few prostate cancer patients in this study to have sufficient with a treatment response for useful comparisons to be made. No predictive utility of the %Δ ADC parameters has been demonstrated independent to tumour group analysis, but a lack of utility has not been proven.

11.3.2 %Δ tDV

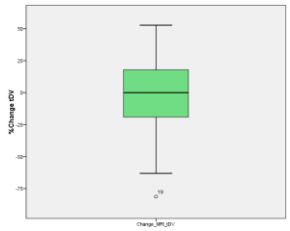
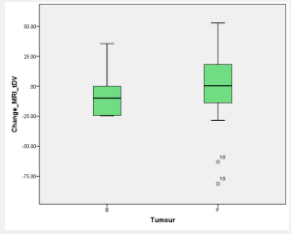
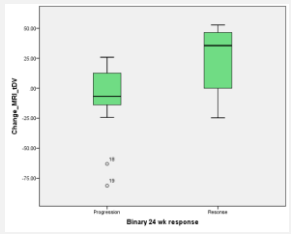
	%Δ tDV
	Mean % Change (min-max) [median; SD]
	-2.84 (-81.29-52.75) [0; 33.81]
	
	Breast **n=5** / **Prostate** **n=14**
	-4.67 (-24.66-35.54) [-9.88; 24.76] / **-2.19** (-81.29-52.75) [0.36; 37.32]
Tumour Groups	
	p=0.893
	PD **n=13** / **Response** **n=5**
	-10.44 (-81.29-25.85) [-6.82; 31.22] / **22.02** (-24.66-52.75) [35.54; 33.12]
Response	
	p=0.216

Table 11-11: DW-MRI Per-Patient Analysis; %Δ tDV; Descriptive statistics and box-plots; All Patients; Tumour groups (B=Breast cancer, P=Prostate Cancer)

Table 11-11 summaries the %ΔtDV data. There no significant difference between the %Δ tDV of breast cancer and prostate cancer patients (independent samples t-test: where t(17)=-0.137, p=0.893).

Patients with a treatment response seem to have shown a larger increase in tDV between scans. This is counterintuitive - a treatment response would be expected to result in smaller lesions. This difference is not statistical significance (where t(16)=-1.288, p=0.216), and may represent a data anomaly, or may suggest influence of the difference between the tumour types. Tumour sub-group response analysis can be reviewed in Table 11-12.

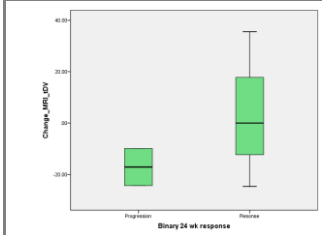
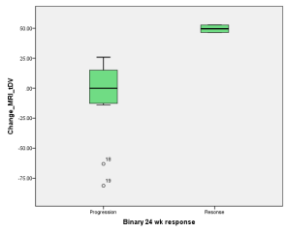
%ΔtDV			
Mean (%) (Min-Max) [Median; SD]			
Breast **n=5**		**Prostate** **n=14**	
PD **n=2**	**Response** **n=3**	**PD** **n=12**	**Response** **n=2**
-17.11 (-24.33--9.88) [-17.11; 10.22]	**3.63** (-24.66-35.54) [0; 30.27]	**25.85** (0.73--6.82) [18.13; 17.63]	**33.89** (0--1.37) [10.22; 1.14]
			
p=1.0		p=0.026	

Table 11-12: DW-MRI Per-Patient Analysis; Response Analysis; %ΔtDV; Tumour Subgroup Analysis

Between the breast patients there is no demonstrable significant difference in the distribution of %ΔtDV between response groups, but with only 5 patients for analysis, statistical comparison is of limited value.

An increase in tDV between scans is predictive of a sustained treatment response for the prostate cancer patients (p=0.026). ROC analysis of only the prostate cancer patients suggests a treatment response can be identified with a sensitivity of 100% and specificity of 100% for lesions with a %ΔtDV> +36.15% (AUC=1.00, p=0.03 using Mann-Whitney U tests for comparison).

This is an unexpected finding, and was also identified with per-lesion analysis. There was significant inter-lesion heterogeneity identified within individual patients during per-lesion analysis, showing some lesions increased in size in patients with overall responsive disease. Healing bone metastases undergo a sclerotic reaction, and sclerotic bone restricts the DW-MRI image further. The sclerotic nature of prostate cancer metastases explains the lower measured ADC parameters of the prostate cancer bone lesions relative to the breast cancer lesions. The high b-value images have been used for VOI definition because of the exaggeration of restricted diffusion, which has been used as a surrogate for identifying the extent of the bone metastases. The sclerotic reaction in bone metastases may extend beyond the true edge of the metastasis, and a treatment response, with healing sclerosis, might result in a larger area of restriction on the high b-value image, and a larger VOI and tDV.

11.3.3 %ΔADC Heterogeneity Parameters

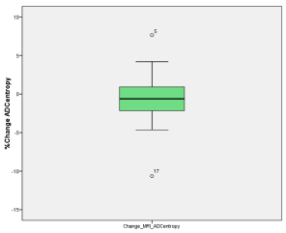
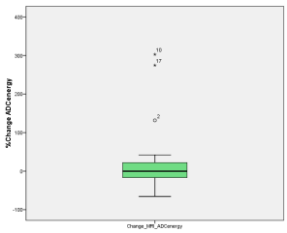
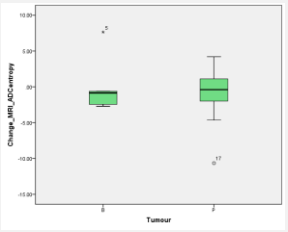
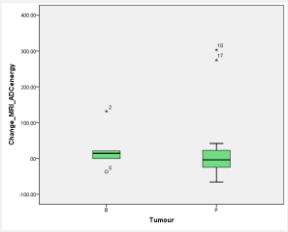
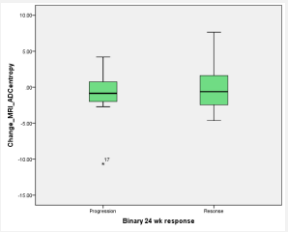
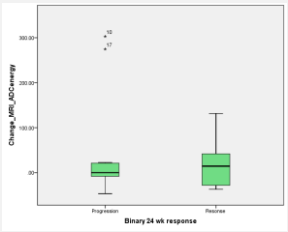
	$\%\Delta ADC_{entropy}$		$\%\Delta ADC_{energy}$	
	Mean % Change (min-max) [median; SD]			
	-0.6 (-10.65-7.64) [-0.62; 3.67]		**31.51** (-65.67-302.89) [-0.09; 99.53]	
				
	Breast **n=5**	**Prostate** **n=14**	**Breast** **n=5**	**Prostate** **n=14**
	0.2 (-2.71-7.64) [-0.85; 4.26]	**-0.88** (-10.65-4.2) [-0.39; 3.56]	**26.01** (-37.03-131.78) [14.51; 63.27]	**33.47** (-65.67-302.89) [-3.74; 111.66]
Tumour Groups				
	p=0.687		p=0.622	
	PD **n=13**	**Response** **n=5**	**PD** **n=13**	**Response** **n=5**
	-0.98 (-10.65-4.2) [-0.85; 3.51]	**0.31** (-4.64-7.64) [-0.62; 4.7]	**-7.38** (4.26-302.89) [11.8; -46.9]	**-46.9** (302.89-111.36) [2.06; -0.09]
				
	p=0.924		p=0.100	

Table 11-13: DW-MRI Per-Patient Analysis; %Δ ADC Heterogeneity Parameters; Descriptive statistics and box-plots; All Patients; Tumour groups (B=Breast cancer, P=Prostate Cancer)

Table 11-13 summarises the data for the %ΔADC Heterogeneity parameters. No statistically significant difference can be identified in the distribution of the %Δ heterogeneity parameters between the tumour groups (Mann-Whitney U tests: $\%\Delta ADC_{entropy}$ p=0.687; $\%\Delta ADC_{energy}$ p=0.622).

There is no statistically significant difference identified in these patients in the distribution of the %Δ heterogeneity parameters between the response groups (Mann Whitney U tests: $\%\Delta ADC_{entropy}$ p=0.924; $\%\Delta ADC_{energy}$ p=0.1). The %Δ ADC heterogeneity parameters do not predict treatment response for these patients. The per-lesion analysis (see previous chapter) suggested $\%\Delta ADC_{energy}$ might have utility for predicting the treatment response. The following

box-plots (Tbl.11-14) display the tumour-specific response analysis for these data, comparing the distributions with Mann-Whitney U tests:

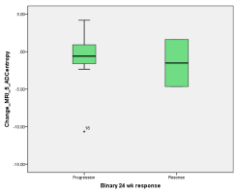
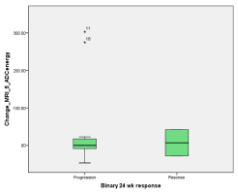
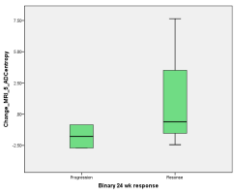
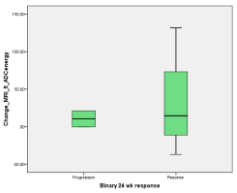
Prostate		Breast	
%ΔADC$_{entropy}$	%ΔADC$_{energy}$	%ΔADC$_{entropy}$	%ΔADC$_{energy}$
			
p=0.923	p=0.923	p=0.4	p=1.0

Table 11-14: DW-MRI Per-Patient Analysis; Response Analysis; %Δ ADC Heterogeneity Parameters; Tumour Subgroup Analysis

There is no statistically significant difference in the distributions between the response groups for either tumour group. However, with such small numbers this is to be expected. The distributions of the parameters for both response group overlap, and there is no clear pattern to suggest an association in these patients.

11.4 %Δ Parameters Correlations

Table 11-15 summarises pairwise correlation analyses between the baseline scan parameters (analysed using Spearman Rank correlation tests).

	%ΔADC$_{mean}$	**%ΔADC$_{median}$**	**%ΔtDV**	**%ΔADC$_{entropy}$**	**%ΔADC$_{energy}$**
%ΔADC$_{mean}$	1	-	-	-	-
%ΔADC$_{median}$	0.96 (<0.001)	1	-	-	-
%ΔtDV	0.24 (0.3287)	0.3 (0.2077)	1	-	-
%ΔADC$_{entropy}$	0.44 (0.0581)	0.43 (0.0687)	-0.07 (0.767)	1	-
%ΔADC$_{energy}$	-0.26 (0.2898)	-0.27 (0.2698)	0.15 (0.5302)	-0.79 (0.0001)	1

Table 11-15: DW-MRI Per-Patient Analysis; %Δ Parameters; Correlations (and p-values) using Spearman Rank analysis; All lesions

The correlations between the parameters are similar to the baseline imaging, and similar to the correlations identified with per-lesion analysis.

11.5 Discussion

		ADC_{mean}	ADC_{median}	tDV	$ADC_{entropy}$	ADC_{energy}
Baseline	**Tumour**	Breast>Prostate p=0.028	Breast>Prostate p=0.058	Prostate>Breast p=0.033	X	Prostate>Breast p=0.06
	Response	X	X	X	X	X
%Δ	**Tumour**	X	X	X	X	X
	Response	X	X	↑ Prostate only p=0.026	X	X

Table 11-16: DW-MRI; Per-Patient Analysis; Response Analysis; summary table

The per-lesion analysis of these data considerable intra-patient inter-lesion heterogeneity. This per-patient analysis attempts to translate the clinical associations of the parameters identified from per-lesion analysis into a clinically relevant per-patient methodology.

The per-lesion analysis did not identify utility of the baseline parameters for predicting the treatment response. The per-patient analysis has also failed to confirm predictive information from these parameters.

A lower baseline ADC has been identified in other soft tissue tumours as a predictor of therapeutic response[213 2016,218,219]. There is no evidence supporting the predictive utility for bone metastases. A lack of association between the baseline ADC and treatment response of bone metastases from prostate cancer has been identified by Messiou et al.[152]. No association has been statistically confirmed in this study between the baseline ADC and the treatment response for either tumour group; the non-statistically significant suggestion of a lower baseline ADC in those with PD is likely to be largely influenced by the lower baseline ADC identified for the prostate patients, possibly due to the more sclerotic nature of these deposits. The prostate cancer patients were much more likely to have PD compared to the breast cancer patients. There are too few data to enable meaningful tumour subgroup analysis. There are demonstrable differences in the ADC parameters of the tumour groups, and predictive utility of the baseline ADC parameters cannot be discounted as this study group has been too small to interrogate fully.

The baseline tDV has no predictive utility in these patients, at either the per-lesion or per-patient level. The volume of osseous disease is prognostic for prostate cancer[222,223], with a smaller volume of disease being associated with a better prognosis. However, there is no evidence showing and association between smaller volume of bone metastases and treatment response.

The baseline ADC heterogeneity parameters have also failed to yield treatment response prognostic value. The per-lesion analysis suggested, for the breast cancer lesions, a lower $ADC_{entropy}$ and higher ADC_{energy} predicted for a treatment response. The per-patient box-plots for breast cancer suggest this association is evident at the patient level, but the differences do not meet statistical significance. There were only 2 breast cancer patients with PD, and therefore meaningful conclusions cannot be drawn from this analysis, but an association between baseline ADC heterogeneity parameters cannot be excluded for the breast cancer patients and warrants further investigation.

The %Δ parameter analysis has also been hampered by the relatively small sample size. There are clear differences identified in the ADC parameters between the tumour types, and drawing conclusions from analysis of all the patients together is likely to be misleading. An example is the %ΔADC parameters – a statistically significant association is identified between the response groups, but it is unlikely this is independent of the tumour subgroup (breast cancer patients had a higher baseline ADC, had a larger increase in ADC between scans, and 60% of the breast cancer patients had a sustained response at 24-weeks). The tumour sub-group analyses do not reveal a significant association between the %Δ ADC and treatment response for either tumour group, but in both tumour groups one of the response groups contains only 2 patients (2 prostate patients had a treatment response and 2 breast patients had PD). In the per-lesion analysis, prostate cancer patients with an increase in ADC between scans was associated with a treatment response. The absence of this relationship at the per-patient level cannot be concluded and needs to be interrogated with a larger study population.

The %Δ tDV behaved in a surprising manner during the per-lesion analysis, showing an increase was associated with a treatment response. For the prostate cancer patients, there is a positive association between an increase in tDV between scans and the likelihood of a treatment response ($p= 0.026$). The inter-lesion heterogeneity was demonstrated with the per-lesion analysis; perhaps the target lesions have not been representative of the clinically relevant disease. Alternatively, the tDV may not represent accurately the size of the tumour deposit within the bone. It is not possible to separate the influence of the tumour and bone stroma signal from the ADC data, but both will inevitably have impact. A treatment response and associated up-regulation of osteoblastic activity and sclerosis may extend beyond the extent of the bone metastasis, causing a halo of increased image restriction around the target volume, but included within the VOI volume on the high b-value image.

The %Δ ADC heterogeneity parameters have shown no utility for prediction of the treatment response. No association was identified either at the per-lesion analysis. In summary, the differences in the parameters between breast and prostate cancer lesions are comparable with those identified at per-lesion analysis, but no clear utility for prediction of treatment response has been identified, apart from an increase in tDV between scans for the prostate cancer patients; this may represent an increased volume of ADC restriction from the healing osteoblastic response in the bone.

In summary, the parameter characteristics are different between breast and prostate cancer metastases, as identified with per-lesion analysis. No parameter has suggested utility for predicting the treatment response apart from an increase in tDV for prostate cancer patients between cans. This suggests that tDV is also measuring the peri-tumour bone architecture and associated changes with therapy. Although per-lesion analysis suggested more predictive utility of the parameters, this has not been seen with per-patient analysis; this may reflect the smaller sample size, or a dilution of the signal due to inter-lesion heterogeneity.

Chapter 12 DW-MRI Per-Patient Analysis – Survival Analysis

The methods of this chapter and the statistical approaches for this section can be reviewed in Chapter 2.3.12.

12.1 OS Analyses

20 patients from this study had data available for OS analysis, 5 breast cancer patients and 15 prostate cancer patients. No breast cancer patients had died at the time of data analysis, and 5 prostate cancer patients had died (33.3%).

12.1.1 Tumour Group Analysis

KM OS analysis of the tumour groups can be seen below (Fig.12-1). There is no statistically significant separation of the survival curves (p=0.154), but certainly a trend towards a significant dichotomisation.

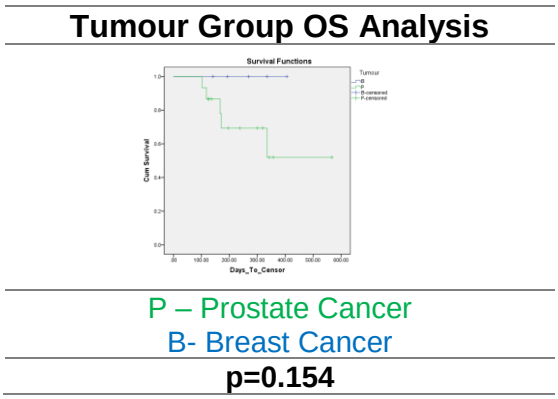


Figure 12-1: DW-MRI Per-Patient Analysis; KM OS Analysis of tumour groups; All Patients; Log Rank Analysis

12.1.2 Baseline Scan OS Analysis

12.1.2.1 ADC Parameters

The Cox regression analysis for the baseline ADC parameters does not demonstrate a significant correlation between the parameters and the OS (Tbl.12-1).

	ADC_{mean}	ADC_{median}
Cox regression p-value [HR (95%CI)]	**0.447** [0.086 (0-47.652)]	**0.595** [0.180 (0.00-101.166)]

Table 12-1: DW-MRI Per-Patient Analysis; Cox regression OS Univariate Analysis; Baseline ADC Parameters

KM OS analysis is recorded in Figure 12-2.

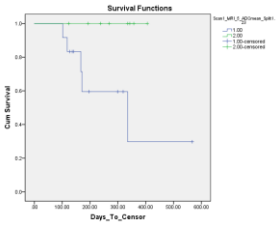
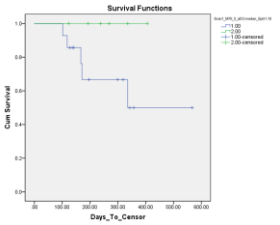
ADC_{mean}	ADC_{median}
	
Dichotomised at 1.23 Low (1) - blue line High (2) - green line	Dichotomised at 1.18 Low (1) - blue line High (2) - green line
p=0.021	**p=0.109**

Figure 12-2: DW-MRI Per-Patient Analysis; KM OS Analysis; Baseline ADC Parameters

For both parameters, the survival curves separate, suggesting a survival benefit for patients with a higher baseline ADC (ADC_{mean}, p=0.021). There were no patients with a baseline ADC_{mean} lower than 1.23 who had died at the time of data analysis. The survival plot for ADC_{median} shows a similar separation, but this has not reached significance (p=0.109), but suggests a trend.

The prostate cancer patients were more likely to have died, and they had a lower baseline ADC. KM OS analysis of the prostate cancer patients can be seen in Figure 12-3.

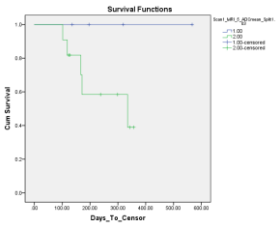
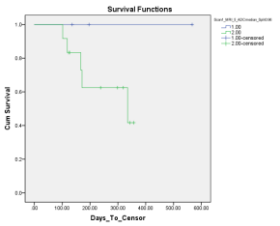
ADC_{mean} (prostate only)	ADC_{median} (prostate only)
	
Dichotomised at 1.03 Low (1) - blue line High (2) - green line	Dichotomised at 0.96 Low (1) - blue line High (2) - green line
p=0.155	**p=0.238**

Figure 12-3: DW-MRI Per-Patient Analysis; KM OS Analysis with Log Rank test for comparison; Baseline ADC Parameters; Prostate patients only

These plots contradict the analysis of the whole patient group, suggesting the tumour subgroup has a larger impact on survival. Although the separation of the curves does not reach statistical significance, the prostate cancer patients with a higher baseline ADC were more likely to have died. The per-lesion analysis also demonstrated this for the prostate cancer lesions. Increased sclerosis of bone metastases has been associated with a survival benefit[235], and increased sclerosis causes results in a lower measured ADC due to restricted water diffusion[234].

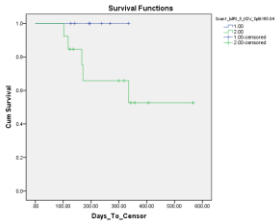
12.1.2.2 tDV

	tDV
Cox regression p-value [HR (95%CI)]	**0.377** [1.003 (0.996-1.01)]

Table 12-2: DW-MRI Per-Patient Analysis; Cox regression OS Univariate Analysis; Baseline tDV

Cox regression analysis of baseline tDV does not show any correlation between the volume of lesion(s) and OS (Tbl.12-2). The data has been dichotomised for KM OS analysis (Fig.12-4):

tDV^{VOI5}



Dichotomised at 165.04cm^3
Low (1) - blue line
High (2) - green line
p=0.115

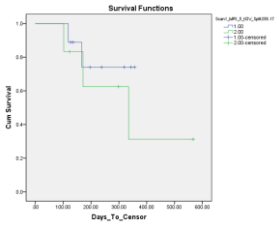
Figure 12-4: DW-MRI Per-Patient Analysis; KM OS Analysis with Log Rank test for comparison; Baseline tDV

There is clear separation of the survival plots, suggesting an OS advantage for patients with smaller volume disease, but this separation does not reach statistical significance (p=0.115), concluding a trend.

The prostate cancer lesions were larger than the breast cancer lesions, and deaths were only in prostate cancer patients; it is not clear that tDV is an independent predictor of OS. KM analysis of the prostate cancer patients (Fig.12-5) does not show significant prognostic utility, implying tDV is not independent of the tumour type. tDV has not demonstrated utility in predicting OS.

tDV

(prostate only)



Dichotomised at 258.17cm^3
Low (1) - blue line
High (2) - green line
p=0.413

Figure 12-5: DW-MRI Per-Patient Analysis; KM OS Analysis with Log Rank test for comparison; Baseline tDV; Prostate patients only

12.1.2.3 Heterogeneity Parameters

	$ADC_{entropy}$	ADC_{energy}
Cox regression p-value [HR (95%CI)]	**0.73** [0.334 (0.001-169.826)]	**0.469** [0 (0-9.61E+104)]

Table 12-3: DW-MRI Per-Patient Analysis; Cox regression OS Univariate Analysis; Baseline ADC Heterogeneity Parameters

Cox regression analysis of the heterogeneity parameters from the baseline imaging do not demonstrate any significant correlation between the heterogeneity parameter and OS (Tbl.12-3).

KM analyses are in Figure 12-6.

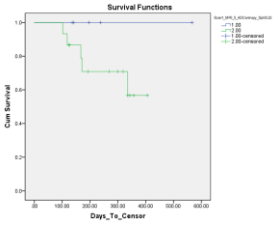
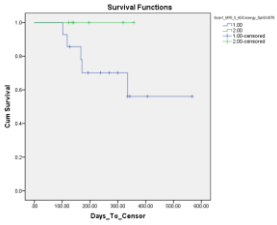
$ADC_{entropy}$	ADC_{energy}
	
Dichotomised at 5.22 Low (1) - blue line High (2) - green line	Dichotomised at 0.0079 Low (1) - blue line High (2) - green line
p=0.227	**p=0.198**

Figure 12-6: DW-MRI Per-Patient Analysis; KM OS Analysis with Log Rank test for comparison; Baseline ADC Heterogeneity Parameters

The patients with higher entropy and/or lower energy (i.e. a wider, more random distribution of ADC voxels and/or a less uniform distribution of ADC voxels) on their pre-treatment DW-MRI seem to have a shorter OS, but this is not proven statistically ($ADC_{entropy}$/ADC_{energy} p=0.227/0.198).

Subgroup analysis of the prostate cancer patients can be seen in Figure 12-7. Prostate cancer patients with a lower baseline $ADC_{entropy}$ and higher baseline ADC_{energy} (i.e. more homogeneous ADC voxel distribution) have an OS benefit, although this has not reached statistical significance, probably because of the sample size ($ADC_{entropy}$ p=0.129; ADC_{energy} p=0.113)

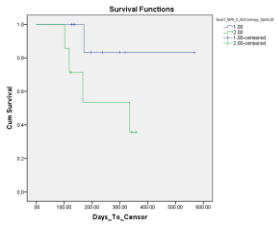
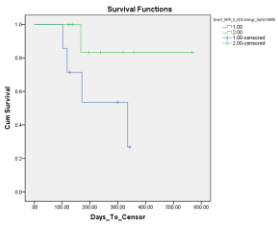
$ADC_{entropy}$ (prostate only)	ADC_{energy} (prostate only)
	
Dichotomised at 5.28 Low (1) - blue line High (2) - green line	Dichotomised at 0.00695 Low (1) - blue line High (2) - green line
p=0.129	p=0.113

Figure 12-7: DW-MRI Per-Patient Analysis; KM OS Analysis with Log Rank test for comparison; Baseline ADC Heterogeneity Parameters; Prostate patients only

12.1.3 %ΔOS Analysis

12.1.3.1 %ΔADC Parameters

	%ΔADC$_{mean}$	%ΔADC$_{median}$
Cox regression p-value [HR (95%CI)]	**0.674** [1.011 (0.96-1.065)]	**0.883** [1.003 (0.962-1.046)]

Table 12-4: DW-MRI Per-Patient Analysis; Cox regression OS Univariate Analysis; %Δ ADC Parameters

There is no demonstrable correlation between the %ΔADC parameters and OS when analysed using Cox regression (Tbl.12-4). The following plots (Fig.12-8) display the KM survival analysis of %ΔADC parameters.

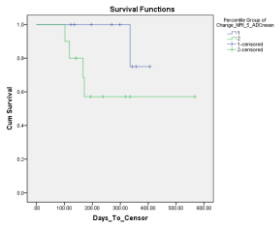
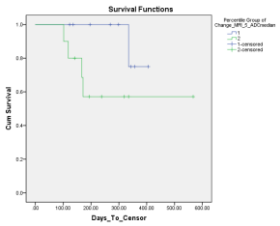
%ΔADC$_{mean}$	%ΔADC$_{median}$
	
Dichotomised at +4.45% (median) Low (1) - blue line High (2) - green line	Dichotomised at +3.36% (median) Low (1) - blue line High (2) - green line
p=0.129	**p=0.129**

Figure 12-8: DW-MRI Per-Patient Analysis; KM OS Analysis with Log Rank test for comparison; %ΔADC Parameters

There is no statistically significant separation of the OS curves for either %ΔADC$_{mean}$ or %ΔADC$_{median}$. There was a significant impact from the tumour group on survival analysis using the ADC parameters from the baseline scan. The plots below (Fig.12-9) record the KM analysis of the prostate cancer patients.

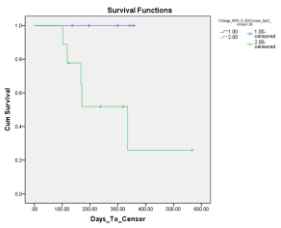
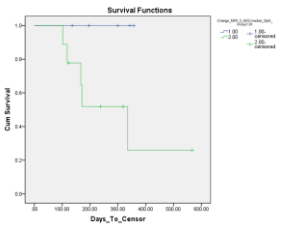
%ΔADC$_{mean}$ (prostate only)	%ΔADC$_{median}$ (prostate only)
	
Dichotomised at -1.30% Low (1) - blue line High (2) - green line	Dichotomised at -1.91% Low (1) - blue line High (2) - green line
p=0.057	**p=0.057**

Figure 12-9: DW-MRI Per-Patient Analysis; KM OS Analysis with Log Rank test for comparison; %Δ ADC Parameters; Prostate patients only

The OS curves separate in the same direction as for the whole-population analysis, but for just the prostate cancer patients it seems the %Δ ADC parameters have prognostic utility for OS. The separations do not quite reach statistical significance at the 0.05 level (p=0.057 for both

parameters), but a trend suggests patients with a larger decrease in ADC between scans have a survival advantage. This confirms the findings from the per-lesion analysis.

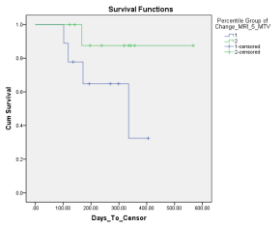
12.1.3.2 %ΔtDV

	%ΔtDV
Cox regression p-value [HR (95%CI)]	**0.353** [0.988 (0.963-1.014)]

Table 12-5: DW-MRI Per-Patient Analysis; Cox regression OS Univariate Analysis; %Δ tDV

Cox regression analysis does not identify significant correlations the %ΔtDV and the OS (Tbl.12-5). Figure 12-10 reports the KM OS analysis for these patients.

%ΔtDV^{VOI5}

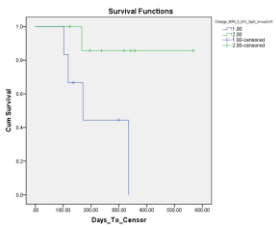


Dichotomised at -7.9% (median)
Low (1) - blue line
High (2) - green line
p=0.091

Figure 12-10: DW-MRI Per-Patient Analysis; KM OS Analysis; %Δ tDV

There is separation of the survival curves; patients with a more positive change or increase of the tDV between scans have a better OS. Neither of these findings reach statistical significance, but there is a trend towards significance (%ΔtDV p=0.091). This is not the anticipated result. Patients with an increase in tumour burden were expected to have a poorer OS. However, throughout this analysis the impact on analysis of the tumour subgroups has been significant. The KM OS analysis of just the prostate cancer lesions can be reviewed in Table 12-11, showing a significant separation of the survival curves, with a survival advantage for the prostate cancer patients with a %ΔtDV>-3.41% (p=0.028). This was not identified with per-lesion analysis.

%ΔtDV
(prostate only)



Dichotomised at -3.41%
Low (1) - blue line
High (2) - green line
p=0.028

Figure 12-11: DW-MRI Per-Patient Analysis; KM OS Analysis; %Δ tDV; Prostate patients only

The tDV has been assumed to be a proxy measure of the tumour within the bone, but this finding suggests the target volumes are also including the surrounding bony architecture influenced by

both the tumour and by the treatment response. Sclerosis causes restriction of the DW-MRI image, and healing bone metastases undergo sclerosis. This sclerotic reaction may extend to include peri-tumour bone. The volumes have been defined using the high b-value images, exaggerating the impact of differences in diffusion between the tissues, and it is possible the volume of image restriction of a metastasis increases when the sclerotic reaction is considered. This was also demonstrated with the response analysis of these patients (see previous chapter) where an increase in tDV predicted a treatment response.

12.1.3.3 %Δ Heterogeneity Parameters

	%ΔADC$_{entropy}$	%ΔADC$_{energy}$
Cox regression p-value [HR (95%CI)]	**0.601** [1.076 (0.818-1.414)]	**0.328** [0.988 (0.966-1.012)]

Table 12-6: DW-MRI Per-Patient Analysis; Cox regression OS Univariate Analysis; %Δ ADC Heterogeneity Parameters

There is no significant correlation between the magnitude of the %ΔADC heterogeneity parameters and OS when using Cox regression analysis (Tbl.12-6).

There is no significant separation of survival curves when using KM analysis - the plots can be seen below (Fig.12-12).

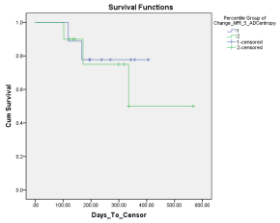
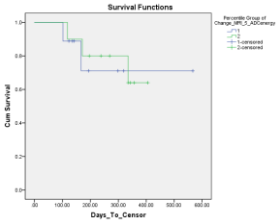
%ΔADC$_{entropy}$	%ΔADC$_{energy}$
	
Dichotomised at -0.62% (median) Low (1) - blue line High (2) - green line	Dichotomised at -0.09% (median) Low (1) - blue line High (2) - green line
p=0.601	**p=0.809**

Figure 12-12: DW-MRI Per-Patient Analysis; KM OS Analysis; %Δ ADC Heterogeneity Parameters

KM analysis of just the prostate cancer patients can be seen in Figure 12-13.

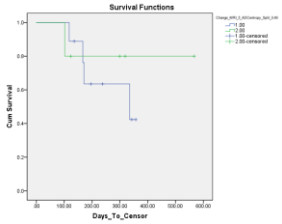
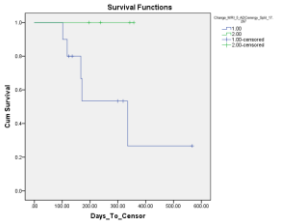
%ΔADC$_{entropy}$ (prostate only)	%ΔADC$_{energy}$ (prostate only)
	
Dichotomised at +0.49% Low (1) - blue line High (2) - green line	Dichotomised at +17.3% Low (1) - blue line High (2) - green line
p=0.595	**p=0.071**

Figure 12-13: DW-MRI Per-Patient Analysis; KM OS Analysis; %Δ ADC Heterogeneity Parameters; Prostate patients only

None of the OS curves reach statistical significance, but there is a trend towards significance for $\%\Delta ADC_{energy}$ (p=0.071) suggesting patients with a greater increase in image uniformity (i.e. a shift towards homogeneity) had a survival benefit. This was identified for the prostate cancer lesions with per-lesion analysis.

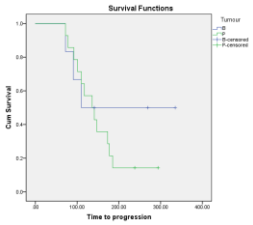
12.2 PFS Analyses

19 patients from this study had data available for PFS analysis, 5 breast cancer patients and 14 prostate cancer patients. Only 2 breast cancer patients had progressed at the time of data analysis (40%), and 12 prostate cancer patients had PD (73.7%). 18 patients had data available using the %Δ parameters (5 breast patients and 13 prostate cancer patients).

12.2.1 Tumour Analysis

KM PFS analysis of these patients dichotomising the data into the tumour groups can be seen below (Fig.12-14). There is no statistically significant separation of the PFS curves (p=0.242). The breast cancer patients have a median PFS of 110 days from treatment initiation, and the prostate cancer patients have a median survival of 136 days (91-180 days 95% CI). The breast patient group, however, contains two patients with much longer PFS (no progression at the time of data analysis).

Tumour Group PFS Analysis



P – Prostate Cancer
B- Breast Cancer
p=0.242

Figure 12-14: DW-MRI Per-Patient Analysis; KM PFS Analysis of tumour groups

12.2.2 Baseline Scan PFS Analysis

12.2.2.1 ADC Parameters

	ADC_{mean}	ADC_{median}
Cox regression p-value **[HR (95%CI)]**	**0.063** [0.034 (0.001-1.2)]	**0.1** [0.038 (0.001-1.859)]

Table 12-7: DW-MRI Per-Patient Analysis; Cox regression PFS Univariate Analysis; Baseline Scan ADC Parameters

Cox analysis (Tbl.12-7) demonstrates a trend towards a significant correlation between ADC_{mean} and the time to PD (ADC_{mean} HR 0.034, p=0.063). A higher ADC_{mean} is correlated with a lower risk of progression. For these patients, there is no significant correlation between the ADC_{median} parameters and time to progression.

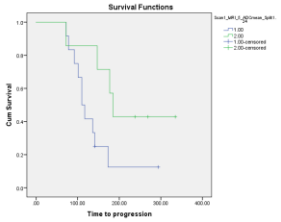
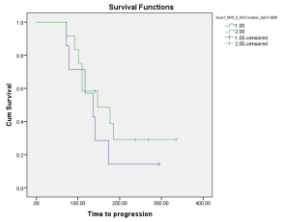
		ADC_{mean}	ADC_{median}
KM plot			
Dichotomisation		1.24	1.0626
Median PFS in days (95% CI)	≤	110 (91.9-128.1)	136 (87.24-184.8)
	>	185 (164.5-205.5)	147 (46.96-247.04)
p-Value (Log Rank)		0.046	0.358

Table 12-8: DW-MRI Per-Patient Analysis; KM PFS Analysis; Baseline ADC Parameters

The KM plots can be reviewed in Table 12-8, corroborating the Cox regression analysis. There is a significant 75 day median PFS benefit for those with an ADC_{mean}>1.24(p=0.046). The ADC_{median} analyses show some separation of the survival curves, again with an OS benefit suggested for patients with higher baseline ADC measurements, but the separation does not meet statistical significance (p=0.358).

For these patients the ADC parameters were lower for prostate cancer lesions and more prostate patients suffered PD. KM analysis of the ADC parameters for the prostate cancer patients can be seen in Table 12-9, showing a PFS benefit for prostate patients with a higher baseline ADC_{mean} (p=0.006).

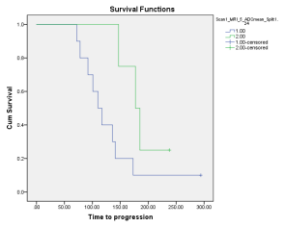
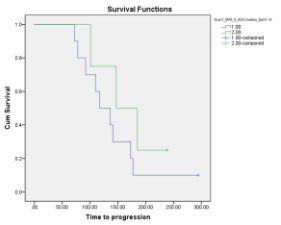
		ADC_{mean} (prostate only)	ADC_{median} (prostate only)
KM plot			
Dichotomisation		1.24	1.14
Median PFS in days (95% CI)	≤	117 (76.7-157.3)	110 (85.2-134.8)
	>	147 (64.7-229.3)	177 (139.8-214.2)
p-Value (Log Rank)		0.08	0.279

Table 12-9: DW-MRI Per-Patient Analysis; KM PFS Analysis; Baseline ADC Parameters; Prostate patients only

12.2.2.2 tDV

	tDV
Cox regression p-value [HR (95%CI)]	**0.055** [1.003 (1-1.007)]

Table 12-10: DW-MRI Per-Patient Analysis; Cox regression PFS Univariate Analysis; Baseline Scan tDV

There is a nearly significant correlation between tDV and PFS in these patients (Cox regression – Tbl.12-10), suggesting a higher baseline tDV is correlated with a higher risk of progression (HR 1.003, p=0.055). KM survival plots with the relevant survival tables can be reviewed in Table 12-11.

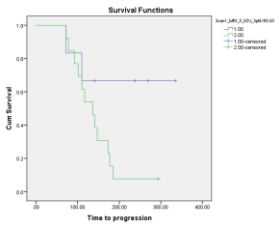
		tDV^{VOI5}
KM plot		
Dichotomisation		165.03cm³ (median)
Median PFS in days (95% CI)	**≤**	n/a
	>	136 (99.6-172.4)
p-Value (Log Rank)		0.064

Table 12-11: DW-MRI Per-Patient Analysis; KM PFS Analysis; Baseline tDV

There is separation of the survival curves, showing a trend towards a PFS benefit for patients with a tDV ≤165.03cm³ (p=0.064).

The prostate cancer patients had larger tDV values. Analysis of the prostate patients (Tbl.12-12) suggests a PFS benefit for smaller lesions, but no statistically significant benefit is demonstrated, possibly due to the small number of patients. The per-lesion analysis did not identify an association between the baseline tDV and PFS for the prostate cancer patients.

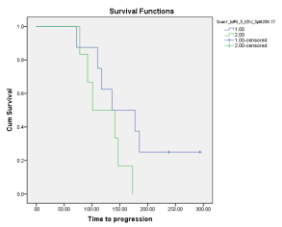
		tDV^{VOI5} (prostate only)
KM plot		
Dichotomisation		258.17cm³
Median PFS in days (95% CI)	**≤**	136 (52.8-219.2)
	>	101 (42.2-159.8)
p-Value (Log Rank)		0.117

Table 12-12: DW-MRI Per-Patient Analysis; KM PFS Analysis; Baseline tDV; Prostate patients only

12.2.2.3 Heterogeneity Parameters

	$ADC_{entropy}$	ADC_{energy}
Cox regression p-value [HR (95%CI)]	**0.822** [1.552 (0.033-71.893)]	**0.832** [0 (0-8.34E+44)]

Table 12-13: DW-MRI Per-Patient Analysis; Cox regression PFS Univariate Analysis; Baseline Scan ADC Heterogeneity Parameters

Cox regression analysis (Tab.12-13) of the baseline heterogeneity parameters does not identify any significant correlation between the magnitude of the parameters and the length of PFS of these patients.

KM PFS plots for the ADC heterogeneity parameters have been reported in Table 12-14. There is no significant separation of the survival curves (p=0.317), but the plot suggests a small PFS advantage for those patients with lower $ADC_{entropy}$ (i.e. more homogeneous lesions). The curves for ADC_{energy} do not separate.

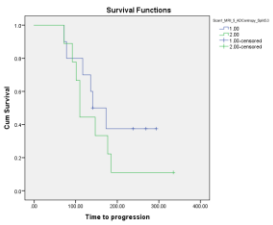
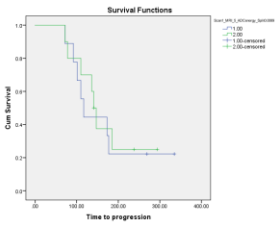
		$ADC_{entropy}$	ADC_{energy}
KM plot			
Dichotomisation		5.3 (median)	0.0069
Median PFS in days (95% CI)	**≤**	141 (90-192)	117 (96.5-137.5)
	>	110 (96.9-123.1)	141 (125.8-156.2)
p-Value (Log Rank)		0.317	0.741

Table 12-14: DW-MRI Per-Patient Analysis; KM PFS Analysis; Baseline ADC heterogeneity parameters

12.2.3 %ΔPFS Analysis

12.2.3.1 %ΔADC Parameters

	%ΔADC$_{mean}$	%ΔADC$_{median}$
Cox regression p-value [HR (95%CI)]	**0.325** [0.981 (0.945-1.019)]	**0.285** [0.983 (0.951-1.015)]

Table 12-15: DW-MRI Per-Patient Analysis; Cox regression PFS Univariate Analysis; %Δ ADC Parameters

Cox regression analysis (Tbl.12-15) does not identify any significant correlation between the magnitude of the ADC$_{mean}$ and ADC$_{median}$ parameters with PFS in these patients. The graphs and survival tables below (Tbl.12-16) demonstrate KM analysis of the ADC data.

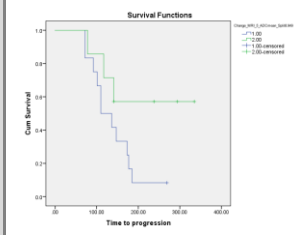
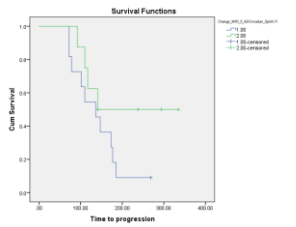
		%ΔADC$_{mean}$	%ΔADC$_{median}$
KM plot			
Dichotomisation		+8.949%	+4.71%
Median PFS in days (95% CI)	≤	110 (70.4-149.6)	136 (86.4-185.6)
	>	n/a	141 (n/a)
p-Value (Log Rank)		0.085	0.152

Table 12-16: DW-MRI Per-Patient Analysis; KM PFS Analysis; %Δ ADC Parameters

There is no demonstrable statistically significant separation of the PFS curves for the %ΔADC parameters, but there is a trend towards significance for %Δ ADC$_{mean}$ (p=0.085), suggesting a larger increase in ADC between the scans is associated with a PFS benefit. However, analysis of just the prostate cancer patients does not demonstrate any PFS predictive utility from the %Δ ADC parameters (Tbl.12-17).

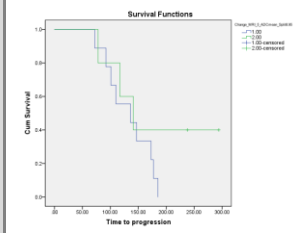
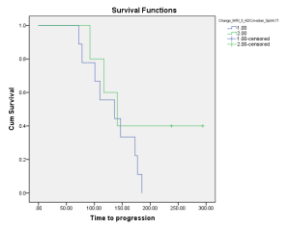
		%ΔADC$_{mean}$ (prostate only)	%ΔADC$_{median}$ (prostate only)
KM plot			
Dichotomisation		8.95	4.71
Median PFS in days (95% CI)	≤	136 (60-212)	136 (60-212)
	>	141 (89.5-192.5)	141 (89.5-192.5)
p-Value (Log Rank)		0.237	0.219

Table 12-17: DW-MRI Per-Patient Analysis; KM PFS Analysis; %ΔADC Parameters; Prostate patients only

12.2.3.2 %Δ Volumetric Parameters

	%ΔtDV
Cox regression p-value [HR (95%CI)]	**0.12** [0.989 (0.975-1.003)]

Table 12-18: DW-MRI Per-Patient Analysis; Cox regression PFS Univariate Analysis; %ΔtDV

There is no correlation identified in these patients between the percentage change in the tDV and the PFS (Cox regression – Tbl.12-18). The KM graphs and survival table below (Tbl.12-19) record the PFs of the study population.

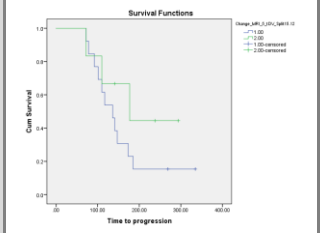
		%ΔtDV
KM plot		
Dichotomisation		+15.12%
Median PFS in days (95% CI)	≤	136 (99.6-172.4)
	>	177 (45.7-308.3)
p-Value (Log Rank)		0.239

Table 12-19: DW-MRI Per-Patient Analysis; KM PFS Analysis; %ΔtDV

There is no significant separation of the KM PFS curves (p=0.239) but there is the suggestion of a PFS with an increase in tDV between scans. This unexpected finding was also identified with OS analysis. The KM PFS analysis of just the prostate cancer patients (Tbl.12-23) confirms this finding, although there is only a trend towards statistical significance (p=0.062). tDV is probably not a measure of the true tumour volume, but is also encompassing the peri-tumour bone changes architectural changes.

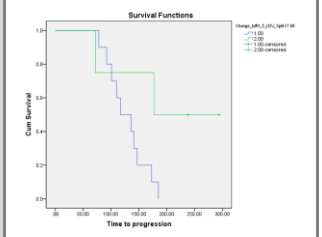
		%ΔtDV (prostate only)
KM plot		
Dichotomisation		17.88
Median PFS in days (95% CI)	≤	117 (76.7-157.3)
	>	177 (0-0)
p-Value (Log Rank)		0.062

Table 12-20: DW-MRI Per-Patient Analysis; KM PFS Analysis; %ΔtDV; Prostate patients only

12.2.3.3 %Δ Heterogeneity Parameters

	%ΔADC$_{entropy}$	%ΔADC$_{energy}$
Cox regression p-value [HR (95%CI)]	**0.941** [1.005 (0.884-1.142)]	**0.917** [1 (0.994-1.006)]

Table 12-21: DW-MRI Per-Patient Analysis; Cox regression PFS Univariate Analysis; %Δ ADC Heterogeneity Parameters

Cox regression analysis (Table 12-21) does not demonstrate a significant correlation between the magnitude of the heterogeneity parameters and the PFS of these patients.

KM analysis for the %ΔADC heterogeneity parameters is recorded in Table 12-22. There is no significant separation of the survival curves. There is perhaps a suggestion of a small PFS benefit for patients with an increase in ADC$_{energy}$ between the scans but a larger patient population would be necessary to interrogate this further.

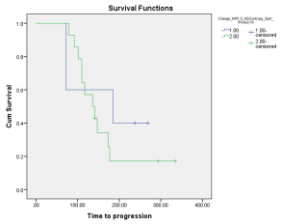
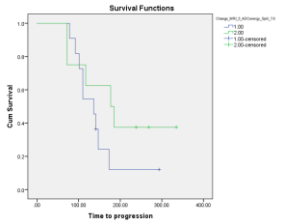
		%ΔADC$_{entropy}$	%ΔADC$_{energy}$
KM plot			
Dichotomisation		-2.16	+7.6% (median)
Median PFS in days (95% CI)	≤	136 (103.6-168.4)	185 (0-427.6)
	>	177 (82.8-271.2)	136 (92-180)
p-Value (Log Rank)		0.449	0.2

Table 12-22: DW-MRI Per-Patient Analysis; KM PFS Analysis; %ΔADC Parameters

The same KM PFS analysis has been performed below for only the prostate cancer patients (Tbl.12-23); no significant separation is identified, but, again, perhaps an increase in ADC$_{energy}$ might be found in patients with a prolonged PFS.

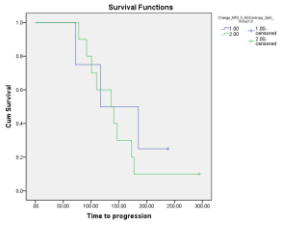
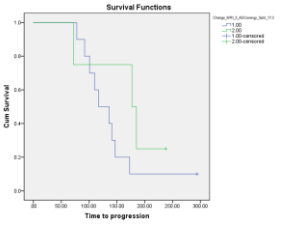
		%ΔADC$_{entropy}$ (prostate only)	%ΔADC$_{energy}$ (prostate only)
KM plot			
Dichotomisation		-1.6	17.3
Median PFS in days (95% CI)	≤	117 (76.7-157.3)	117 (6.3-227.7)
	>	177 (66.3-287.7)	136 (88-184)
p-Value (Log Rank)		0.486	0.215

Table 12-23: DW-MRI Per-Patient Analysis; KM PFS Analysis; %Δ ADC heterogeneity parameters; Prostate patients only

12.3 Discussion

		ADC_{mean}	ADC_{median}	tDV	$ADC_{entropy}$	ADC_{energy}
Baseline	↑OS	↓ PCa only p=0.155	↓ PCa only p=0.238	X	↓ All pts p=0.227 PCa p=0.129	↑ All pts p=0.198 PCa p=0.113
	↑PFS	↑ All pts p=0.002 PCa p=0.006	↑ All pts p=0.056 PCa p=0.094	↓ PCa p=0.138	↓ p=0.022	↑ p=0.021
%Δ	↑OS	↓ PCa only p=0.057	↓ PCa only p=0.057	↑ Both VOI All pts p=0.096 PCa p=0.028	X	↑ PCa only p=0.071
	↑PFS	X	X	↑ All pts p=0.027 PCa p=0.085	X	↑ All p=0.05 PCa p=0.078

Table 12-24: DW-MRI Per-Patient Analysis; Survival analyses; Summary table

The previous chapter demonstrated little predictive utility from the DW-MRI in these patients. However, there appears to be more useful information for prognosis.

The statistical analysis of the OS and PFS for the two tumour groups does not identify a statistically significant difference, but this is probably misleading; only one breast cancer patient had died at the time of data analysis, and only 2 of the breast patients developed PD. With so few events in the study group it is difficult to draw firm conclusions about the OS/PFS of each of the tumour groups in this study.

The baseline OS analysis of the ADC parameters is difficult to interpret. When analysing all the patients together the ADC parameters predict OS, with a benefit identified for patients with a higher baseline ADC. However, analysis of the prostate patients only shows an OS benefit for those with a lower baseline ADC; ADC parameters are not independent to the tumour type. This OS benefit for prostate cancer lesions was also demonstrated with the per-lesion. There is no direct evidence in the literature to support or refute this finding. Increased sclerosis causes more diffusion restriction and a lower ADC[234]. There is an association between increased sclerosis of bone metastases and a better prognosis[235] in breast cancer metastases.

The %Δ in ADC parameters have no utility in these patients for predicting OS over and above the differentiation of the tumour sub-groups. With the per-lesion analysis there was a trend of an association between a fall in ADC and a PFS benefit for the prostate cancer lesions. This effect has not been translated through to per-patient analysis.

The baseline tDV data suggests an OS benefit from smaller lesions, but the impact of the tumour types is likely to be significant, and analysis of the prostate patients only is non-significant. The PFS analysis also only suggests a trend towards a PFS benefit for the prostate cancer patients, but this again is not significant. A greater burden of bone metastases in prostate cancer is a marker of shorter OS[240], and the development of the bone scan index as a measure of total skeletal burden indicate a worse survival with a higher burden of disease[222,223,242]. The per-lesion analysis of these patients did not demonstrate a significant association between baseline tDV and PFS for the prostate cancer patients.

The %Δ tDV parameter for OS and PFS analysis is more difficult to justify. tDV is used as a measure of disease burden within the bone, but it is likely, certainly with the methodology utilised in this study, peri-tumour bone is included in the volume, and this is impacted by both the tumour and changes in bone architecture with treatment. An increase in tDV between the scans is associated with an OS and PFS benefit. This is still identified when just prostate cancer patients are interrogated. The per-lesion analysis also demonstrated this finding for the breast cancer lesions. tDV probably encompasses treatment changes in peri-tumour bone architecture.

Increased tumour heterogeneity was anticipated to be associated with a poorer prognosis[172,176]. The per-lesion analysis demonstrated an OS benefit was associated with a lower $ADC_{entropy}$ and higher ADC_{energy} for the prostate cancer lesions, and a higher baseline ADC_{energy} was associated with a PFS benefit for the breast cancer lesions. There is no evidence in the literature reporting how the ADC heterogeneity parameters of bone metastases behave following treatment.

The per-patient analyses show a trend for prognostic utility of the ADC heterogeneity parameters. A lower $ADC_{entropy}$ and high ADC_{energy} are associated with an OS benefit for the whole patient population, and this relationship persists when prostate cancer patients alone are analysed. There is a demonstrable PFS benefit using the baseline ADC heterogeneity parameters; a lower $ADC_{entropy}$ ($p=0.129$) and higher ADC_{energy} ($p=0.113$) are associated with this benefit.

The %ΔADC heterogeneity parameters do not show a strong association with OS, but there is a trend towards an increase in ADC_{energy} being associated with a OS benefit ($p=0.071$). An increase in ADC_{energy} between scans was also predicted a median 75-day PFS benefit with per-lesion analysis, and this association also persists through to analysis of the prostate cancer patients only. This suggests a prognostic benefit for patients with tumours with a more homogeneous ADC voxel distribution before treatment, and those becoming more homogeneous with treatment.

The previous chapter has not identified predictive utility from the baseline or %ΔADC heterogeneity parameters.

In summary, lower baseline ADC measurements in prostate cancer patients indicate an OS and PFS benefit. The ADC parameters are not independent of the tumour groups; breast and prostate cancer quantification parameters behave differently following treatment. Per-patient analysis does not elicit the prognostic utility identified with per-lesion analysis. This may be because of inter-lesion heterogeneity diluting the signal, or the smaller sample size.

Chapter 13 DW-MRI - Whole-Body Quantification of Bone Metastases

The methods of this chapter and the statistical approaches for this section can be reviewed in Chapter 2.3.5. This chapter is presented in 4 sections: 13.1-Descriptive statistics and response analysis; 13.2-OS Analysis; 13.3-PFS Analysis; 13.4-Discussion.

13.1 Descriptive Statistics and Response Analysis

13.1.1 Baseline Scan

13.1.1.1 ADC^{WB} Parameters

The baseline ADC^{WB} Parameters are summarised in Table 13-1. There is no significant difference in the distribution of ADC_{mean}^{WB} and ADC_{median}^{WB} (Paired samples t-test: where ($t(19)=1.287$, $p=0.214$). The two parameters are closely correlated (Pearson correlation: $r=0.998$, $p<0.0001$). A correlation would be expected as these two parameters are describing the distribution of the same voxels in this study group, and the per-lesion and per-patient target lesion analysis also demonstrate a close correlation between these two parameters.

The box-plots show the means of ADC_{WB} parameter distributions are higher for breast cancer patients, but these differences are not statistically significant (non-parametric Mann-Whitney U tests: ADC_{mean}^{WB} $p=0.672$; ADC_{median}^{WB} $p=0.735$).

The per-lesion and per-patient target lesion analyses demonstrate the baseline ADC parameters were lower for the prostate cancer patients. Perhaps with a larger study group this would have been identified with whole-body analysis, but the whole-body parameters are less sensitive for differentiation between the tumour groups using ADC.

The box-plots of treatment response have similar appearances between the baseline ADC parameters. Those with a treatment response perhaps had higher baseline ADC_{WB} measurements, but these differences are not statistically significant (Mann-Whitney U tests: ADC_{mean}^{WB} $p=0.849$; ADC_{median}^{WB} $p=0.924$). Analysis of the prostate cancer lesions only does not demonstrate any significant difference in the ADC parameters between the response groups (ADC_{mean}^{WB} $p=0.513$; ADC_{median}^{WB} $p=0.513$). The baseline ADC parameters from the per-lesion and per-patient target lesion analyses also demonstrated no utility in predicting the treatment response.

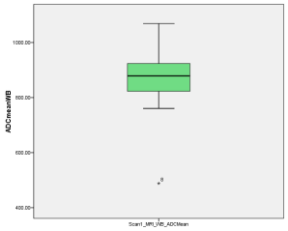
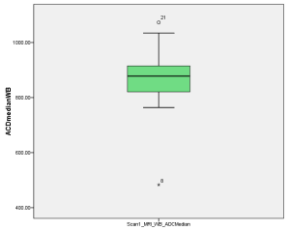
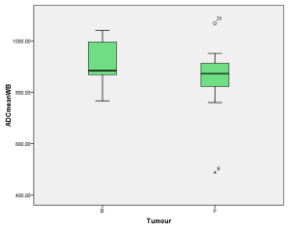
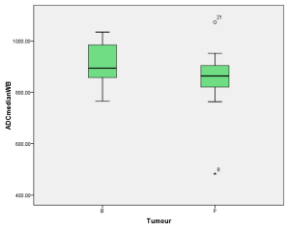
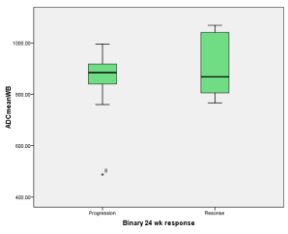
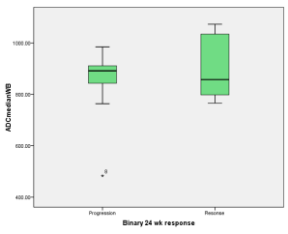
	ADC_{mean}^{WB}	ADC_{median}^{WB}
	mean (10^{-6} mm^2/s) (min-max) [median; SD]	
All patient n=20	**868.7** (488.24-1068.75) [878.95; 122.29]	**866.07** (483.22-1072.95) [877.81; 121.51]
All Patients		
Breast Cancer n=5	**911.32** (766.48-1040.96) [884.78; 108.95]	**907.34** (765.82-1034.4) [894.33; 105.82]
Prostate Cancer n=15	**854.49** (488.24-1068.75) [873.12; 126.65]	**852.31** (483.22-1072.95) [864.07; 126.6]
Tumour Subgroups B=Breast; P=Prostate	 p=0.672	 p=0.735
Progression	**850.27** (488.24-995.9) [884.78; 127.41]	**848.86** (483.22-984.62) [891.55; 125.76]
Response	**910.06** (766.48-1068.75) [868.49; 137.44]	**905.73** (765.82-1072.95) [857.54; 139.67]
Response Groups	 p=0.849	 p=0.924

Table 13-1: WB DW-MRI; Baseline Scan; WB ADC Parameters; All Patients; Descriptive statistics (ADC expressed as 10^{-6} mm^2/s)

13.1.1.2 tDV_{WB}

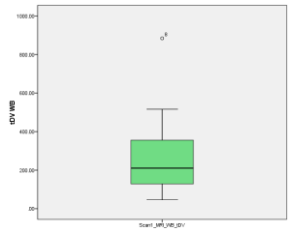
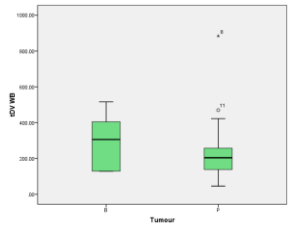
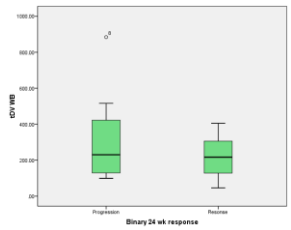
	tDV_{WB} (cm^3)	Box-plots
	Mean (min-max) [median; SD]	
All patient n=20	**266.86** (45.92-884.04) [210.41; 194.84]	
Breast Cancer (B) n=5	**296.95** (128.1-516.05) [306.1; 170.59]	
Prostate Cancer (P) n=15	**256.83** (45.92-884.04) [204.24; 206.82]	p=0.553
Progression n=	**302.11** (99.24-884.04) [230.19; 221.42]	
Response n=	**220.38** (45.92-405.17) [216.58; 141.83]	p=0.410

Table 13-2: WB DW-MRI; Baseline Scan; WB tDV; All Patients; Descriptive statistics

The box-plots (Tbl.13-2) suggest the breast patients had a larger volume of tDV_{WB} identified, but this is not statistically significant (Mann-Whitney U test: p=0.553).

The data also suggests larger tDV_{WB} measurements were more likely to be associated with PD, but again this difference is not significant (Mann-Whitney U test: p=0.633). When analysing the prostate patients independently, there is no demonstrable difference in the tDV_{WB} distribution between the response groups (p=0.410).

The baseline tDV_{WB} does not predict treatment response for these patients.

13.1.1.3 ADC_{WB} Heterogeneity Parameters

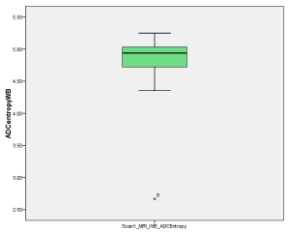
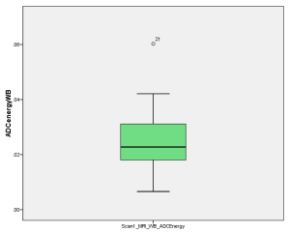
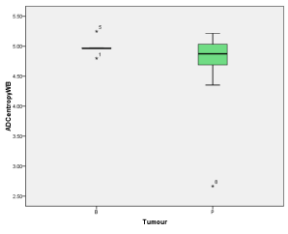
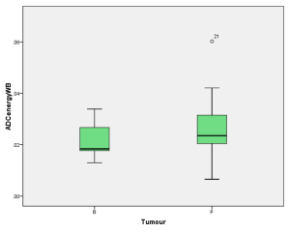
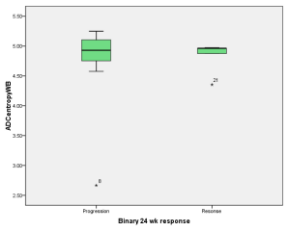
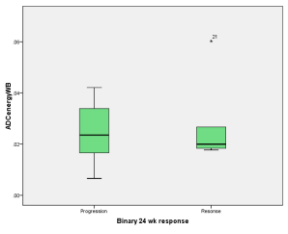
	$ADC_{entropy}^{WB}$	ADC_{energy}^{WB}
	mean (min-max) [median; SD]	
All patient n=20	**4.79** (2.67-5.25) [4.94; 0.54]	**0.0259** (0.0066-0.0603) [0.0228; 0.0121]
All Patients		
Breast Cancer n=5	**4.99** (4.8-5.25) [4.96; 0.16]	**0.0219** (0.013-0.0339) [0.0184; 0.0083]
Prostate Cancer n=15	**4.72** (2.67-5.22) [4.87; 0.61]	**0.0273** (0.0066-0.0603) [0.0235; 0.0131]
Tumour Subgroups B=Breast; P=Prostate	 p=0.197	 p=0.230
Progression	**4.76** (2.67-5.25) [4.93; 0.66]	**0.0251** (0.0066-0.0421) [0.0235; 0.0109]
Response	**4.83** (4.35-4.97) [4.96; 0.27]	**0.0286** (0.0177-0.0603) [0.02; 0.018]
Response Groups	 p=0.173	 p=0.633

Table 13-3: WB DW-MRI; Baseline Scan; WB ADC Heterogeneity Parameters; All Patients; Descriptive statistics

The box-plots (Tbl.13-3) suggest the prostate cancer patients have a lower $ADC_{entropy}^{WB}$ and a higher ADC_{energy}^{WB} than the breast cancer patients, suggesting the whole-body parameters are tending towards more homogeneity for the prostate cancer patients. However, these differences are not statistically significant (Mann-Whitney U tests: $ADC_{entropy}^{WB}$ p=0.197, ADC_{energy}^{WB} p=0.230).

There is no demonstrable difference in the distribution of the ADC_{WB} parameters between the response groups (Mann-Whitney U tests: $ADC_{entropy}^{WB,}$ p=0.173, ADC_{energy}^{WB} p=0.633).

Sub-group analysis of the tumour groups also shows no significant difference in the between the response groups (Prostate: $ADC_{entropy}^{WB}$ p=0.103; ADC_{energy}^{WB} p=0.410; Breast: $ADC_{entropy}^{WB}$ p=0.2; ADC_{energy}^{WB} p=0.1).

Baseline Parameters Correlations

The table below (Tbl.13-4) summarises pairwise correlation analysis between the baseline scan parameters (analysed using Spearman Rank correlation tests).

	ADC_{mean}^{WB}	ADC_{median}^{WB}	tDV^{WB}	$ADC_{entropy}^{WB}$	ADC_{energy}^{WB}
ADC_{mean}^{WB}	1	-	-	-	-
ADC_{median}^{WB}	0.99 (<0.0001)	1	-	-	-
tDV^{WB}	-0.01 (0.9749)	-0.06 (0.806)	1	-	-
$ADC_{entropy}^{WB}$	-0.18 (0.4426)	-0.1 (0.6865)	-0.75 (0.0001)	1	-
ADC_{energy}^{WB}	0.05 (0.8256)	-0.02 (0.9448)	0.68 (0.0009)	-0.89 (<0.0001)	1

Table 13-4: DW-MRI Per-lesion Analysis; Baseline Parameters; Correlations (and p-values) using Spearman Rank analysis; All lesions

A strong correlation is identified between ADC_{mean} and ADC_{median} parameters, as would be anticipated as they are describing the average of the same VOIs.

There is a strong negative correlation between tDV and $ADC_{entropy}^{WB}$ (correlation coefficient -0.75, p=0.0001), and a strong correlation between tDV and ADC_{energy}^{WB} (correlation coefficient 0.68 p=0.0009).

There is a strong negative correlation between ADC_{energy}^{WB} and $ADC_{entropy}^{WB}$ (correlation coefficient -0.89, p=<0.0001).

13.1.2 Second Scan

n=19

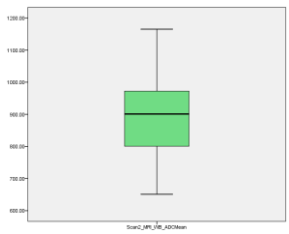
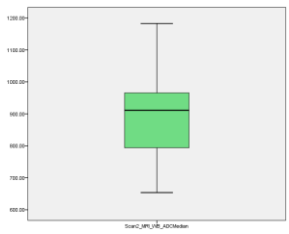
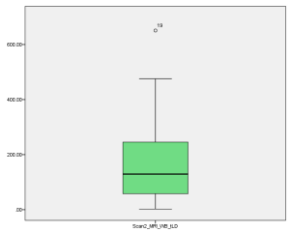
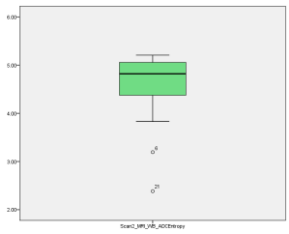
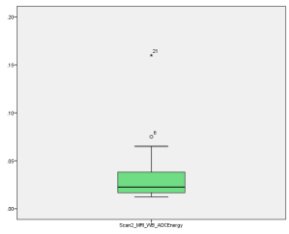
	Mean (min-max) [median; SD]	
ADC_{meanWB} (10^{-6} mm^2/s)	**885.8** (651.18-1165.21) [901.09; 138.98]	
$ADC_{medianWB}$ (10^{-6} mm^2/s)	**886.18** (653.70-1182.51) [911.32; 140.81]	
tDV_{WB} **(cm3)**	**187.43** (2.31-650.89) [129.23; 184.45]	
$ADC_{entropyWB}$	**4.56** (2.38-5.21) [4.82; 0.76]	
$ADC_{energyWB}$	**0.0354** (0.0128-0.5052) [0.0249; 0.1104]	

Table 13-5: WB DW-MRI; Second Scan; WB Parameters; All Patients; Descriptive statistics

13.1.3 Percentage Change (%Δ) Between Scans

13.1.3.1 %Δ ADC^{WB} Parameters

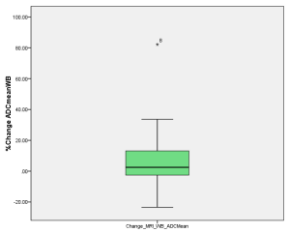
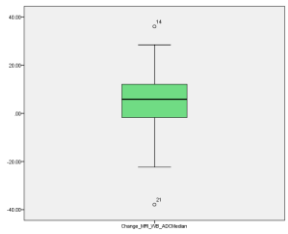
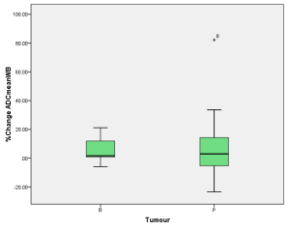
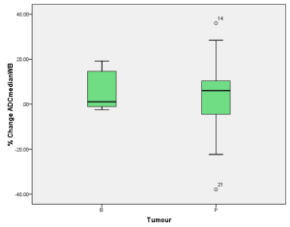
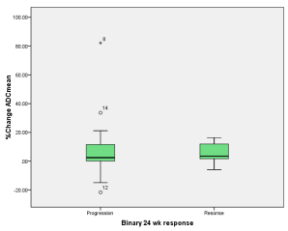
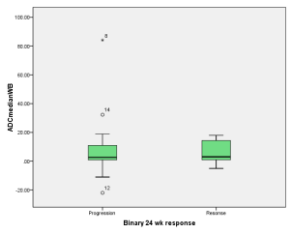
	%ΔADC$_{mean}^{WB}$	%ΔADC$_{median}^{WB}$
	mean (%) (min-max) [median; SD]	
All patient n=19	**7.23** (-23.42-82.15) [2.47; 22.91]	**3.72** (-37.9-36.05) [5.85; 17.15]
All Patients		
Breast Cancer n=5	**5.95** (-5.88-21.15) [1.68; 10.62]	**6.14** (-5.04-18.89) [1.57; 10.02]
Prostate Cancer n=15	**7.69** (-23.42-82.15) [3; 26.3]	**-7.81** (-115.24-84.14) [2.49; 51.84]
Tumour Subgroups B=Breast; P=Prostate	 p=1.0	 p=0.964
Progression	**9.37** (-21.63-82.15) [2.47; 25.98]	**9.49** (-22-84.14) [2.64; 26.03]
Response	**5.52** (-5.88-16.33) [3.52; 8.76]	**6.25** (-5.04-17.95) [3.04; 9.59]
Response Groups	 p=0.387	 p=0.387

Table 13-6: WB DW-MRI; %Δ WB ADC Parameters; All Patients; Descriptive statistics

Table 13-6 summarises the data for the %Δ whole-body ADC parameters. There is no statistically significant difference in the distributions between the response groups (Mann-Whitney U tests: $\%\Delta ADC_{mean}^{WB}$ p=0.387; $\%\Delta ADC_{median}^{WB}$ p=0.387). There is no predictive utility identified when only the prostate cancer patients are analysed (p=0.513 for both parameters).

The per-lesion analysis showed for the prostate cancer patents an increase in ADC_{mean} between scans could predict a treatment response. This has not been shown with the whole-body quantification.

An examination of the waterfall plots (Figs.13-1/13-2) of the individual patient data suggests the changes in whole-body ADC_{mean} and ADC_{median} have not been discriminatory in these patients for identifying a treatment response. Most of the patients had a measured increase in the parameter, irrespective of the response group. The largest fall in the ADC parameters was in the PD group, but this group also had the patient with the largest measured increase.

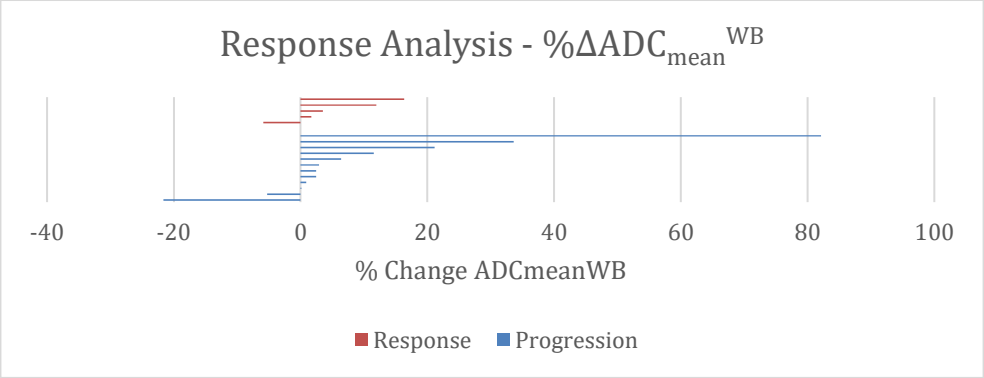


Figure 13-1: WB DW-MRI; %Δ ADC_{meanWB} Parameters; Response Analysis; Waterfall plot of all patients

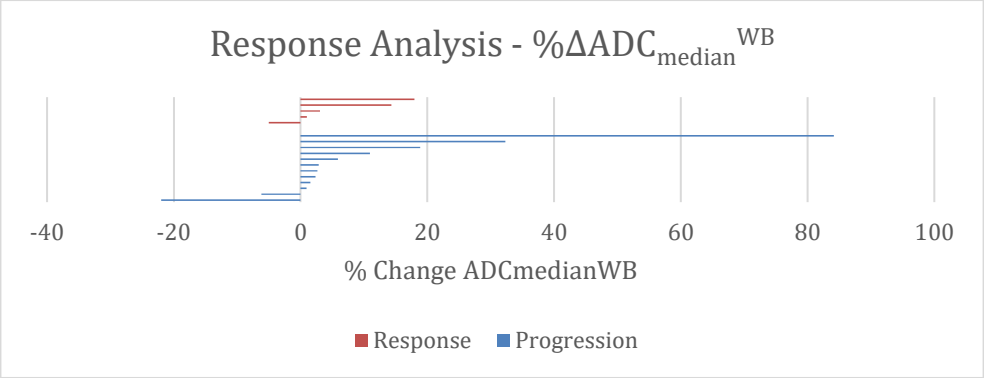


Figure 13-2: WB DW-MRI; %Δ $ADC_{medianWB}$ Parameters; Response Analysis; Waterfall plot of all patients

13.1.3.2 %Δ tDV_{WB}

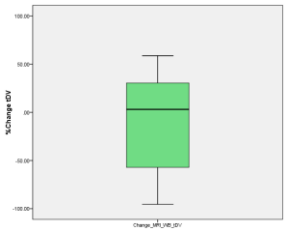
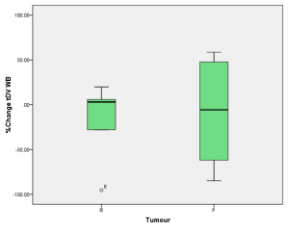
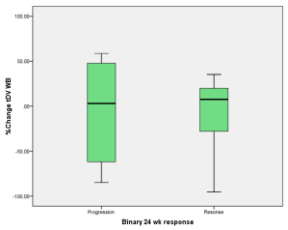
	%ΔtDV_{WB}	Box-plots
	Mean (min-max) [median; SD]	
All patient n=19	**-11.48** (-95.47-58.72) [3.08; 52.14]	
Breast Cancer (B) n=5	**-18.9** (-95.47-19.92) [3.08; 46.22]	
Prostate Cancer (P) n=14	**-8.84** (-84.88-58.72) [-5.61; 55.48]	p=0.823
Progression	**-10.68** (-84.88-58.72) [3.08; 56.27]	
Response	**-12.1** (-95.47-35.3) [7.65; 52.11]	p=0.289

Table 13-7: WB DW-MRI; %Δ WB tDV; All Patients; Descriptive statistics

A wide range of %Δ tDV values have been found in these patients (Tbl.13-7). The range of responses is wider for the prostate cancer patients. The breast cancer patients mostly had a decrease in size of tDV, and the breast cancer patients were more likely to have a clinical response. There is no significant difference in the distribution of %ΔtDV between the tumour groups (Mann-Whitney U test: p=0.823).

Within the progression group there were some lesions with increases and other with decreases in tDV. In the response group, most patients had a decrease in tDV, which would be more in keeping with a pathophysiological response of a tumour deposit. The differences in the distributions between the response groups does not meet statistical significance (Independent samples t-test, where t(16)=0.049, p=0.289). When analysing the prostate cancer patients, no predictive utility for response is shown (p=0.769).

The per-lesion and per-patient target lesion analysis suggested an increase in tDV was associated with a treatment response, probably because of a sclerotic reaction in healing metastases. The whole-body parameter has not demonstrated utility in identifying a treatment response for these patients.

The waterfall plot of the individual data (Fig.13-3) shows the change in tDV in these patients appears to be unhelpful in identifying a treatment response – there are patients with both increase and decreases in tDV in both response groups.

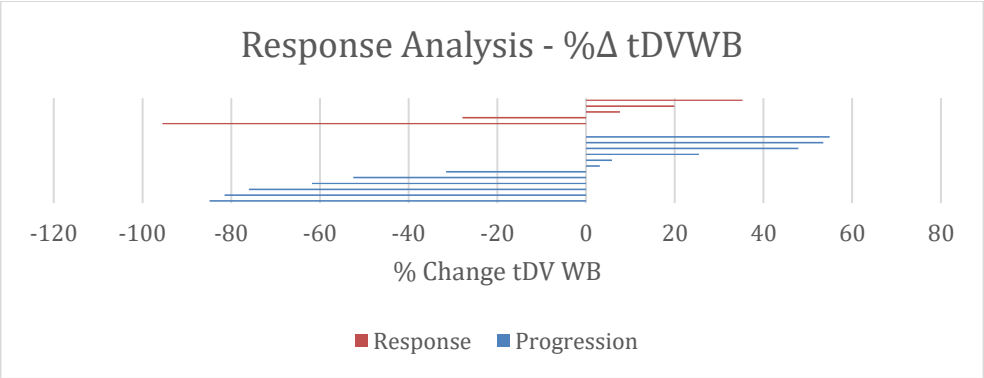


Figure 13-3: WB DW-MRI; %Δ tDV_{WB} Parameters; Response Analysis; Waterfall plot of all patients

13.1.3.3 %Δ ADC^{WB} Heterogeneity Parameters

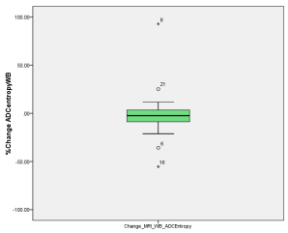
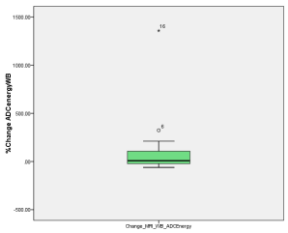
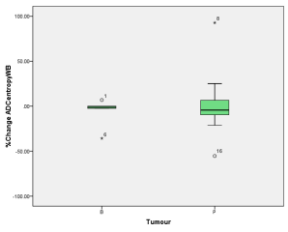
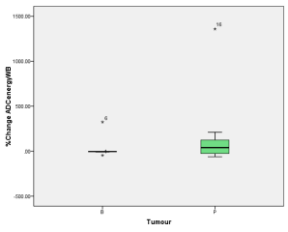
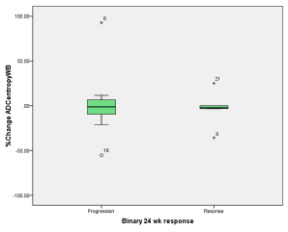
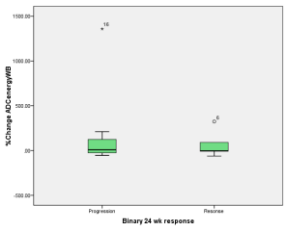
	%ΔADC$_{entropy}^{WB}$	%ΔADC$_{energy}^{WB}$
	mean (min-max) [median; SD]	
All patient n=19	**-0.88** (-55.33-92.83) [-2.36; 28.51]	**112.34** (-61.65-1357.91) [8.03; 317.69]
All Patients		
Breast Cancer n=5	**-6.47** (-35.78-6.92) [-1.18; 16.78]	**53.12** (-45.62-324.04) [-3.07; 152.52]
Prostate Cancer n=15	**1.11** (-55.33-92.83) [-4.28; 31.98]	**133.49** (-61.65-1357.91) [39.34; 361.61]
Tumour Subgroups B=Breast; P=Prostate	 p=1.0	 p=0.823
Progression	**0.36** (-55.33-92.83) [-1.18; 32.52]	**136.47** (-53.95-1357.91) [8.03; 376.05]
Response	**-3.21** (-35.78-25.18) [-2.36; 21.67]	**68.44** (-61.65-324.04) [-3.07; 153.04]
Response Groups	 p=0.246	 p=0.246

Table 13-8: WB DW-MRI; %Δ WB ADC Heterogeneity Parameters; All Patients; Descriptive statistics

Table 13-8 reports the data for the baseline ADC heterogeneity parameters. The baseline imaging suggested the breast cancer patients had a smaller range of ADC$_{entropy}^{WB}$. The box-plots show only small changes in ADC$_{entropy}^{WB}$ were noted for the breast patients following treatment. There is no statistically significant difference in the distributions of %ΔADC$_{entropy}^{WB}$ between the tumour groups (Mann-Whitney U test: p=1.00). The box plot suggests a larger range of change in parameter recorded for the prostate cancer patients, perhaps with a trend towards a larger decrease in entropy (i.e. towards homogeneity).

Similarly, the distributions of $\%\Delta ADC_{energy}^{WB}$ suggest there was little demonstrable change in the whole-body measurement of ADC_{energy}^{WB} for the breast cancer patients. More change (in both directions) was identified for the prostate cancer patients. However, for these data there is demonstrate no statistically significant difference in the distributions between tumour groups (p=0.823). The prostate patients perhaps had more of an increase in ADC_{energy}^{WB}, i.e. a move towards uniformity of ADC voxel distribution.

There is no significant difference between the response groups for either $\%\Delta ADC^{WB}$ response parameter (Mann-Whitney U tests: $\%\Delta ADC_{entropy}^{WB}$ p=0.246, $\%\Delta ADC_{entropy}^{WB}$ p=0.246). Analysis of the prostate cancer patients alone does not identify a utility for predicting the treatment response ($\%\Delta ADC_{entropy}^{WB}$ p=0.308, $\%\Delta ADC_{entropy}^{WB}$ p=0.769).

The waterfall plot below (Figs.13-4/13-5) demonstrates the range of changes within the response groups. Within both response groups there were patients with increases and decreases in the heterogeneity parameters, confirming no utility has been identified for detecting a treatment response.

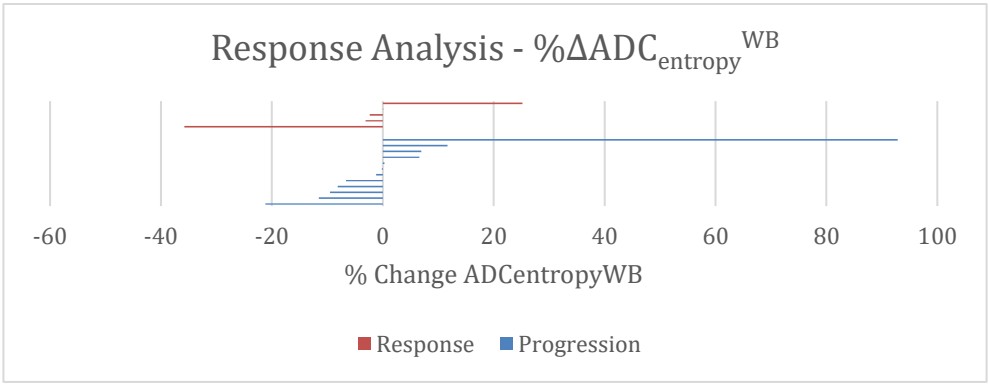


Figure 13-4: WB DW-MRI; %Δ $ADC_{entropy}^{WB}$ Parameters; Response Analysis; Waterfall plot of all patients

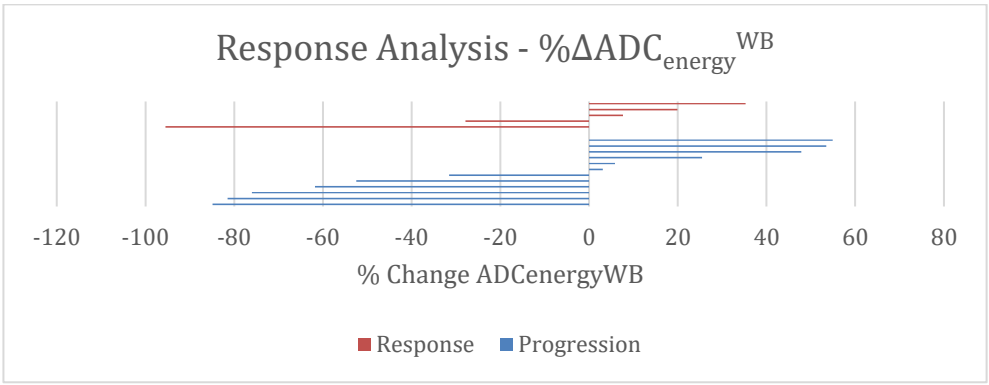


Figure 13-5: WB DW-MRI; %Δ ADC_{energy}^{WB} Parameters; Response Analysis; Waterfall plot of all patients

13.1.3.4 %Δ Parameters Correlations

The table below (Tbl.13-9) summarises pairwise correlation analysis between the baseline scan parameters (analysed using Spearman Rank correlation tests).

	$\%\Delta ADC_{mean}^{WB}$	$\%\Delta ADC_{median}^{WB}$	$\%\Delta tDV^{WB}$	$\%\Delta ADC_{entropy}^{WB}$	$\%\Delta ADC_{energy}^{WB}$
$\%\Delta ADC_{mean}^{WB}$	1	-	-	-	-
$\%\Delta ADC_{median}^{WB}$	0.99 (<0.0001)	1	-	-	-
$\%\Delta tDV^{WB}$	0.27 (0.2665)	0.25 (0.3037)	1	-	-
$\%\Delta ADC_{entropy}^{WB}$	-0.35 (0.1408)	-0.34 (0.1585)	-0.93 (<0.0001)	1	-
$\%\Delta ADC_{energy}^{WB}$	0.24 (0.3144)	0.23 (0.3364)	0.88 (<0.0001)	-0.92 (<0.0001)	1

Table 13-9: DW-MRI Per-lesion Analysis; Baseline Parameters; Correlations (and p-values) using Spearman Rank analysis; All lesions

A strong correlation is identified between $\%\Delta ADC_{mean}$ and $\%\Delta ADC_{median}$ parameters, as would be anticipated as they are describing the average of the same VOIs.

There is a strong negative correlation between %ΔtDV and $\%\Delta ADC_{entropy}^{WB}$ (correlation coefficient -0.93, p<0.0001), and a strong correlation between %ΔtDV and $\%\Delta ADC_{energy}^{WB}$ (correlation coefficient 0.88 p<0.0001).

There is a strong negative correlation between $\%\Delta ADC_{energy}^{WB}$ and $\%\Delta ADC_{entropy}^{WB}$ (correlation coefficient -0.90, p=<0.0001).

13.2 OS Analysis

13.2.1 Tumour Group

20 patients from this study had data available for OS analysis, 5 breast cancer patients and 15 prostate cancer patients. No breast cancer patients had died at the time of data analysis, and 5 prostate cancer patients had died (33.3%).

Chapter 12.1 demonstrated the trend for an OS benefit for the breast cancer patients (KM plot can be seen in Fig.13-6).

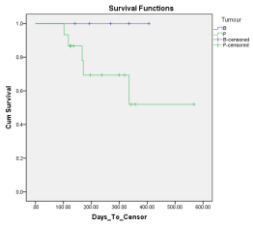
Tumour Group OS Analysis

P – Prostate Cancer B- Breast Cancer
p=0.154

Figure 13-6: DW-MRI Per-Patient Analysis; KM OS Analysis of tumour groups; All Patients; Log Rank Analysis

13.2.2 Baseline Scan

13.2.2.1 ADC_{WB} Parameters

	ADC_{mean}^{WB}	ADC_{median}^{WB}
Cox regression p-value [HR (95%CI)]	**0.610** [1.002 (0.994-1.009)]	**0.611** [1.002 (0.99-1.009)]

Table 13-10: WB DW-MRI; OS Analysis; Baseline Scan; WB ADC Parameters; All Patients; Cox regression analysis

The Cox regression analysis (Tbl.13-10) does not identify any correlation between the baseline ADC^{WB} parameters and the risk of death for these patients.

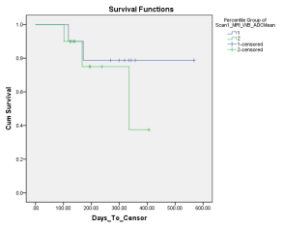
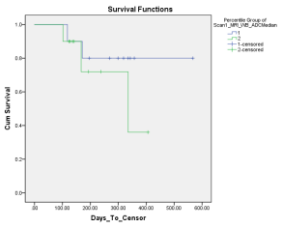
ADC_{mean}^{WB}	ADC_{median}
	
Dichotomised at 878.95 Low (1) - blue line High (2) - green line	Dichotomised at 877.81 Low (1) - blue line High (2) - green line
p=0.422	p=0.341

Figure 13-7: WB DW-MRI; OS Analysis; Baseline Scan; WB ADC Parameters; All Patients; KM Analysis (ADC expressed as 10^{-6} mm^2/s)

Log rank analysis of the KM survival curves (Fig.13-7) does not identify any statistically significant separation for either ADC_{mean}^{WB} or ADC_{median}^{WB}. Although this whole-body analysis has not demonstrated a difference in the ADC parameter distributions between the tumour groups, the per-lesion level showed a lower ADC for the prostate lesions compared to the breast cancer lesions. KM analysis of only the prostate cancer lesions shows a trend towards an OS benefit for the patients with a lower baseline ADC; this trend is particularly strong with ADC_{median}^{WB} (p=0.064, see Fig.13-8).

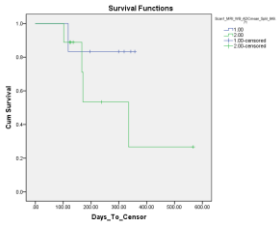
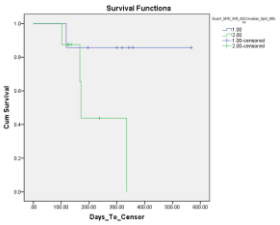
ADC_{mean}^{WB} (Prostate only)	ADC_{median}^{WB} (prostate only)
	
Dichotomised at 849.71 Low (1) - blue line High (2) - green line	Dichotomised at 859.14 Low (1) - blue line High (2) - green line
p=0.212	p=0.064

Figure 13-8: WB DW-MRI; OS Analysis; Baseline Scan; WB ADC Parameters; Prostate patients only; KM Analysis (ADC expressed as 10^{-6} mm^2/s)

13.2.2.2 tDV_{WB}

	tDV_{WB}
Cox regression p-value [HR (95%CI)]	**0.823** [1.00 (0.995-1.004)]

Table 13-11: WB DW-MRI; OS Analysis; Baseline Scan; WB tDV; All Patients; Cox regression analysis

Cox regression analysis (Tbl.13-11) does not identify a significant correlation between tDV_{WB} and the risk of death. The KM analysis below (Fig.13-19) does not demonstrate significant separation of the survival curves (p=0.134). There is however the suggestion of a survival advantage for patients with smaller tDV_{WB}.

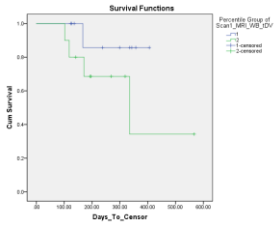
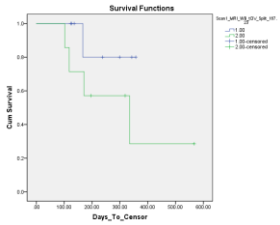
tDV_{WB}	tDV_{WB} (prostate only)
	
Dichotomised at 210.41cm^3 (median) Low (1) - blue line High (2) - green line	Dichotomised at 167.23cm^3 Low (1) - blue line High (2) - green line
p=0.134	p=0.164

Figure 13-9: WB DW-MRI; OS Analysis; Baseline Scan; tDV; KM Analysis

Analysis of only the prostate cancer patients also suggest an OS advantage for the patients with a smaller tDV (Fig.13-9), but again the separation of the survival curves does not reach statistical significance (p=0.164).

13.2.2.3 WB Heterogeneity Parameters

	$ADC_{entropyWB}$	$ADC_{energyWB}$
Cox regression p-value [HR (95%CI)]	**0.507** [2.677 (0.146-48.964)]	**0.745** [0 (0-1.097+26)]

Table 13-12: WB DW-MRI; OS Analysis; Baseline Scan; WB ADC Heterogeneity Parameters; All Patients; Cox regression analysis

Cox regression analysis (Fig.13-12) does not identify a correlation between the baseline whole-body heterogeneity parameters and OS. KM analysis for all the patients can be reviewed below (Fig.3-10):

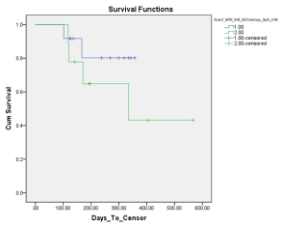
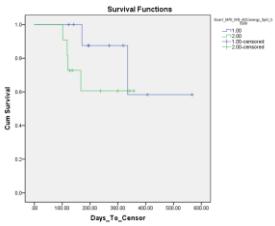
$ADC_{entropyWB}$	$ADC_{energyWB}$
	
Dichotomised at 4.99 Low (1) - blue line High (2) - green line	Dichotomised at 0.0208 Low (1) - blue line High (2) - green line
p=0.281	p=0.328

Figure 13-10: WB DW-MRI; OS Analysis; Baseline Scan; WB ADC Heterogeneity Parameters; All Patients; KM Analysis

There is no significant separation of the survival curves for either parameter when comparing using log-rank analysis. The data has also been analysed for the prostate cancer patients only, but still no statistically significant separation of the PFS survival curves is identified (Fig.13-11).

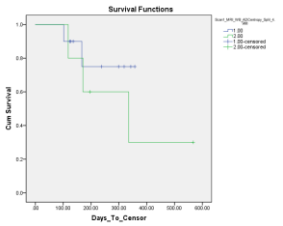
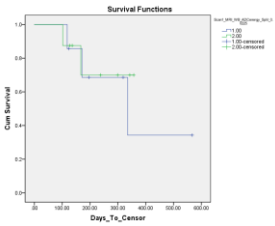
$ADC_{entropyWB}$ (prostate only)	$ADC_{energyWB}$ (prostate only)
	
Dichotomised at 4.988 Low (1) - blue line High (2) - green line	Dichotomised at 0.0225 Low (1) - blue line High (2) - green line
p=0.376	p=0.328

Figure 13-11: WB DW-MRI; OS Analysis; Baseline Scan; WB ADC Heterogeneity Parameters; Prostate patients only; KM Analysis

13.2.3 %Δ Parameters

n=19, 5 deaths (all prostate patients)

13.2.3.1 %ΔADC$_{WB}$ Parameters

	%ΔADC$_{meanWB}$	%ΔADC$_{medianWB}$
Cox regression p-value [HR (95%CI)]	**0.701** [0.991 (0.947-1.037)]	**0.724** [0.992 (0.949-1.037)]

Table 13-13: WB DW-MRI; OS Analysis; %Δ WB ADC Parameters; All Patients; Cox regression analysis

There is no correlation identified in these patients between the size of change of the ADC parameters and OS analysis when compared with Cox regression analysis (Tbl.13-13). KM analysis is reported in Figure 13-12.

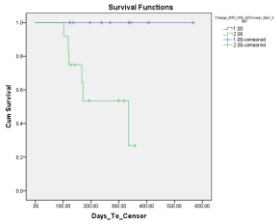
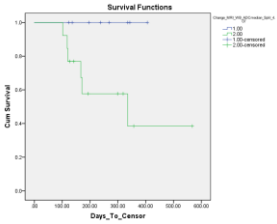
%ΔADC$_{meanWB}$	%ΔADC$_{medianWB}$
	
Dichotomised at +3.9% Low (1) - blue line High (2) - green line	Dichotomised at +4.07% Low (1) - blue line High (2) - green line
p=0.011	p=0.032

Figure 13-12: WB DW-MRI; OS Analysis; %Δ WB ADC Parameters; All Patients; KM Analysis

Significant separation of the OS curves has been identified using the %Δ ADCWB parameters suggesting OS benefit for patients with a more negative, or decrease, in the ADC parameters between the scans. There was no demonstrable difference in the %Δ ADC parameters between the tumour groups for these patients, but subgroup analysis of the prostate cancer patients shows these survival curve separations persist to show OS benefit for a fall in ADC, although with fewer patients for analysis the statistical comparisons are weaker (Fig.13-13).

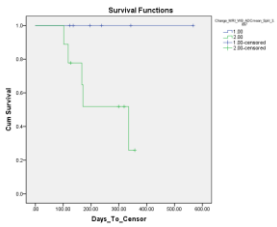
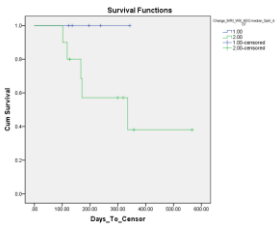
%ΔADC$_{meanWB}$ (prostate only)	%ΔADC$_{medianWB}$ (prostate only)
	
Dichotomised at +3.9% Low (1) - blue line High (2) - green line	Dichotomised at +4.07% Low (1) - blue line High (2) - green line
p=0.049	p=0.128

Figure 13-13: WB DW-MRI; OS Analysis; %Δ WB ADC Parameters; All Patients; KM Analysis

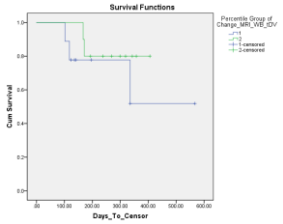
13.2.3.2 %ΔtDV$_{WB}$

	%ΔtDV$_{WB}$
Cox regression p-value [HR (95%CI)]	**0.553** [0.995 (0.977-1.013)]

Table 13-14: WB DW-MRI; OS Analysis; %Δ WB tDV; All Patients; Cox regression analysis

There is no demonstrable correlation in these patients between the change in tDV$_{WB}$ and OS analysis when Cox regression analysis is used (Tbl.13-14). KM analysis of the data can be reviewed below (Fig.13-14).

%tDV$_{WB}$



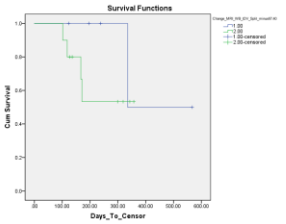
Dichotomised at -31.145%
Low (1) - blue line
High (2) - green line

p=0.371

Figure 13-14: WB DW-MRI; OS Analysis; %Δ WB tDV; All Patients; KM Analysis

There is no significant separation of the survival curves using %ΔtDV$_{WB}$ (p=0.371). Analysis of only the prostate cancer patients (Fig.13-15) also demonstrates no utility for these patients for predicting their OS.

%tDV$_{WB}$ (prostate only)



Dichotomised at -67.4%
Low (1) - blue line
High (2) - green line
p=0.343

Figure 13-15: WB DW-MRI; OS Analysis; %Δ WB tDV; Prostate patients only; KM Analysis

13.2.3.3 %Δ WB Heterogeneity Parameters

	%ΔADC$_{entropyWB}$	%ΔADC$_{energyWB}$
Cox regression p-value [HR (95%CI)]	**0.045** [0.957 (0.917-0.999)]	**0.093** [1.002 (1.00-1.003)]

Table 13-15: WB DW-MRI; OS Analysis; %Δ WB ADC Heterogeneity Parameters; All Patients; Cox regression analysis

There is a demonstrable correlation between the size of change of the whole-body heterogeneity parameters and OS, compared using Cox regression analysis (Tbl.13-15). This is particularly noted for %ΔADC$_{entropyWB}$. A HR of 0.957 (p=0.045) suggests an increase in ADC$_{entropyWB}$ between images is associated with a lower risk of death. For ADC$_{energyWB}$, a fall in the parameter between

scans is associated with an OS benefit (HR 1.002 (p=0.093; which is approaching statistical significance). The data has been split at the median for KM OS analysis (Fig.13-16).

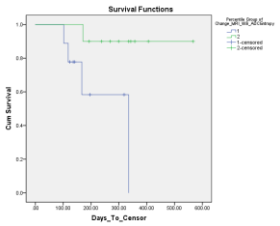
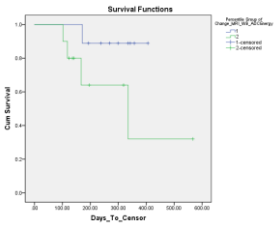
$\%\Delta ADC_{entropyWB}$	$\%\Delta ADC_{energyWB}$
	
Dichotomised at -2.36% Low (1) - blue line High (2) - green line	Dichotomised at +8.03% Low (1) - blue line High (2) - green line
p=0.017	p=0.067

Figure 13-16: WB DW-MRI; OS Analysis; %Δ WB ADC Heterogeneity Parameters; All Patients; KM Analysis

For these patients, there has been a significant separation of the KM OS curves using both %Δ ADC_{WB} Heterogeneity parameters. An OS advantage has been noted for patients with a more positive change in $ADC_{entropy}$ (a move towards heterogeneity), and a more negative change in ADC_{energy} (away from image uniformity). This confirms the Cox regression results.

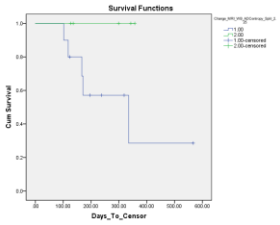
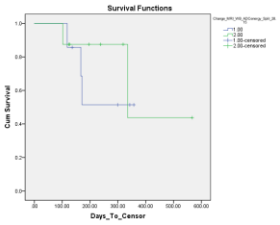
$\%\Delta ADC_{entropyWB}$ (prostate only)	$\%\Delta ADC_{energyWB}$ (prostate only)
	
Dichotomised at +2.25% Low (1) - blue line High (2) - green line	Dichotomised at +28.15% Low (1) - blue line High (2) - green line
p=0.087	p=0.598

Figure 13-17: WB DW-MRI; OS Analysis; %Δ WB ADC Heterogeneity Parameters; Prostate patients only; KM Analysis

The KM plots (Fig.13-17) show the analysis for just the prostate cancer patients. There was no demonstrable difference identified in these patients between the %Δ ADC parameters between the tumour groups, and subgroup analysis limits the statistical validation with a small study group. This analysis suggests persistence of the OS benefit for prostate cancer patients with an increase in $ADC_{entropy}^{WB}$ between the scans, but not for $\%\Delta ADC_{energy}^{WB}$. The %Δ ADC heterogeneity parameters are strongly correlated with the %ΔtDV but not the other parameters measured. %ΔtDV did not show OS predictive utility. Multivariate analysis is not feasible due to the small number of patients to examine if the %Δ heterogeneity parameters are independent predictors of OS.

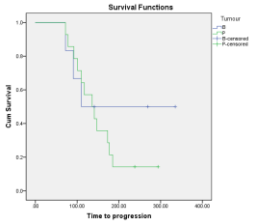
13.3 PFS Analysis

19 patients from this study had data available for PFS analysis, 5 breast cancer patients and 14 prostate cancer patients. Only 2 breast cancer patients had progressed at the time of data analysis (40%), and 12 prostate cancer patients had PD (73.7%). 18 patients had data available using the %Δ parameters (5 breast patients and 13 prostate cancer patients).

13.3.1 Tumour Group

The PFS analysis of the tumour groups has been previously recorded (Chapter 12.2.1). KM PFS analysis shows no significant separation of the PFS curves (p=0.242), but the survival curves show a trend for PFS benefit for the breast patients (Fig.13-18).

Tumour Group PFS Analysis



P – Prostate Cancer
B- Breast Cancer
p=0.242

Figure 13-18: DW-MRI Whole-Body Analysis; KM PFS Analysis of tumour groups

13.3.2 Baseline WB Scan

13.3.2.1 WB ADC_{WB} Parameters

	ADC_{mean}^{WB}	ADC_{median}^{WB}
Cox regression p-value [HR (95%CI)]	**0.941** [1 (0.996-1.004)]	**0.949** [1 (0.996-1.004)]

Table 13-16: WB DW-MRI; PFS Analysis; Baseline Scan; WB ADC Parameters; All Patients; Cox regression analysis

There is no correlation in these patients between the size of the baseline ADC_{WB} parameters and PFS (Cox regression analysis – Tbl.13-16). KM analysis can be reviewed in Table 13-17. There is no significant separation of the PFS survival curves for either baseline ADC_{WB} parameter noted on the KM PFS plots. Analysis of just the prostate cancer lesions (Fig.13-19) also demonstrates no significance. There is no demonstrable difference in the baseline whole-body ADC parameters identified for these but at the per-lesion analysis prostate cancer metastases had a lower ADC than the breast cancer lesions. The breast cancer patients had a PFS benefit.

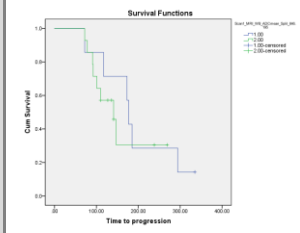
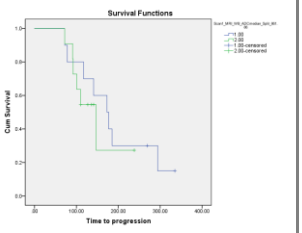
		ADC_{mean}^{WB}	ADC_{median}^{WB}
KM plot			
Dichotomisation		878.95	861.86
Median PFS in days (95% CI)	≤	177 (166.7-187.3)	173 (117.2-228.8)
	>	141 (101-181)	147 (95.7-198.3)
p-Value (Log Rank)		0.557	0.576

Table 13-17: WB DW-MRI; PFS Analysis; Baseline Scan; WB ADC Parameters; All Patients; KM Analysis & Survival Table (ADC expressed as 10^{-6} mm^2/s)

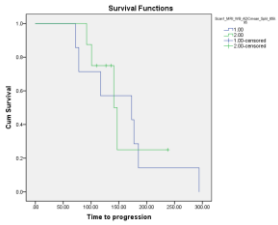
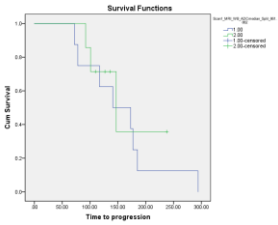
ADC_{mean}^{WB} (prostate only)	ADC_{median}^{WB} (prostate only)
	
Dichotomised at 846.195 Low (1) - blue line High (2) - green line	Dichotomised at 861.86 Low (1) - blue line High (2) - green line
p=0.817	p=0.589

Figure 13-19: WB DW-MRI; PFS Analysis; Baseline Scan; WB ADC Parameters; All Patients; KM Analysis (ADC expressed as 10^{-6} mm^2/s)

13.3.2.2 tDV

	tDV_{WB}
Cox regression p-value [HR (95%CI)]	**0.3** [1.001 (0.999-1.004)

Table 13-18: WB DW-MRI; PFS Analysis; Baseline Scan; WB tDV; All Patients; Cox regression analysis

Cox regression analysis (Tbl.13-18) does not identify a significant correlation between the baseline tDV_{WB} and the ultimate PFS of these patients. The data has been dichotomised for KM analysis (Tbl.13-19). The PFS plots for all patients analysed together suggest a PFS benefit for the smaller skeletal burden of disease (p=0.066). There is no significant difference in these patients between the tDV of the tumour types, but at the per-lesion analysis the prostate cancer lesions were on average larger than the breast cancer lesions, and the prostate cancer patients had a worse PFS. Subgroup analysis of the prostate cancer lesions is therefore appropriate. The separation of the PFS curves persists when only the prostate cancer lesions are analysed (p=0.036).

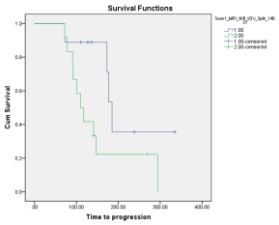
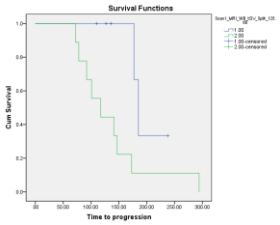
		tDV_{WB}	tDV_{WB} (prostate only)
KM plot			
Dichotomisation		148.31cm³	125.68cm³
Median PFS in days (95% CI)	≤	185 (171.8-198.2)	185 (172.2-197.8)
	>	110 (82.8-137.2)	117 (70.3-163.7)
p-Value (Log Rank)		0.066	0.036

Table 13-19: WB DW-MRI; PFS Analysis; Baseline Scan; WB tDV; KM Analysis Survival Table

13.3.2.3 Heterogeneity Parameters

	$ADC_{entropyWB}$	$ADC_{energyWB}$
Cox regression p-value [HR (95%CI)]	**0.809** [1.148 (0.374-3.519)]	**0.301** [0 (0-1719597023.266)]

Table 13-20: WB DW-MRI; PFS Analysis; Baseline Scan; WB ADC Heterogeneity Parameters; All Patients; Cox regression analysis

Cox regression analysis (Tbl.13-20) demonstrates no significant correlation between the baseline ADC_{WB} heterogeneity parameters and risk of progression. The data has also been analysed with KM PFS analysis. The plots and survival table can be seen below (Tbl.13-21):

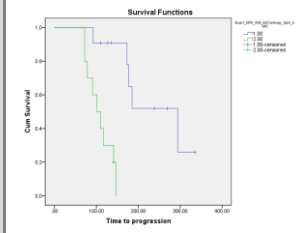
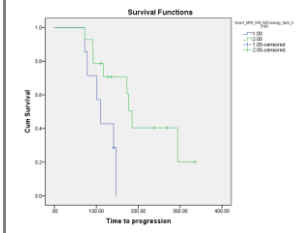
		$ADC_{entropy}^{WB}$	ADC_{energy}^{WB}
KM plot			
Dichotomisation		4.945	0.0183
Median PFS in days (95% CI)	≤	294 (174-414)	110 (86.9-133.1)
	>	101 (71.6-130.4)	185 (167.5-202.5)
p-Value (Log Rank)		0.0002	0.022

Table 13-21: WB DW-MRI; PFS Analysis; Baseline Scan; WB ADC Heterogeneity Parameters; All Patients; KM Analysis Survival Table

KM PFS analysis of the baseline $ADC_{entropy}WB$ parameters for all patients shows a clear and significant separation of the survival curves, with a median 193-day PFS benefit for patients with more homogeneous parameters (lower $ADC_{entropy}^{WB}$ and higher ADC_{energy}^{WB}) (p=0.0002 and 0.022 respectively). There was no significant difference in the distribution of the ADC heterogeneity parameters identified at the whole-body analysis, but analysis of the prostate cancer patients only

shows persistence of this association (Tbl.13-22, p=0.017 for $ADC_{entropy}^{WB}$, and 0.116 for ADC_{energy}^{WB} suggesting a trend towards significance).

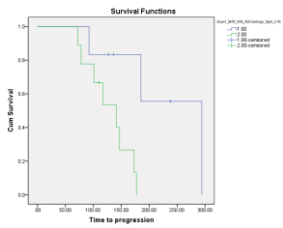
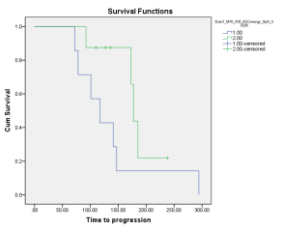
		$ADC_{entropy}^{WB}$ (prostate only)	**ADC_{energy}^{WB} (prostate only)**
KM plot			
Dichotomisation		4.76	0.0226
Median PFS in days (95% CI)	≤	294 (0-0)	117 (75.9-158.1)
	>	141 (89.9-192.1)	177 (168.9-185.1)
p-Value (Log Rank)		0.017	0.116

Table 13-22: WB DW-MRI; PFS Analysis; Baseline Scan; WB ADC Heterogeneity Parameters; Prostate patients only; KM Analysis Survival Table

Correlation assessment has shown strong correlations between tDV_{WB} and the ADC heterogeneity parameters; multivariate analysis to test whether the ADC heterogeneity parameters are independent predictors of OS is not possible due to the small number of patients included.

13.3.3 %Δ Parameters

13.3.3.1 %Δ ADC_{WB} Parameters

	%ΔADC_{meanWB}	**%Δ$ADC_{medianWB}$**
Cox regression p-value [HR (95%CI)]	0.987 [1 (0.976-1.025)]	0.996 [1 (0.976-1.025)]

Table 13-23: WB DW-MRI; PFS Analysis; %Δ WB ADC Parameters; All Patients; Cox regression analysis

There is no statistically significant correlation identified in these patients between the size of the change in the ADC_{WB} parameters and the patients' PFS (Cox regression – Tbl.13-23). The data has been dichotomised at the median for KM analysis (Tbl.13-24). There is a trend towards a significant separation for the %ΔADC_{WB} parameters, with the suggestion on a 63-day median PFS benefit in patients with an increase in the ADC_{WB} parameters between scans.

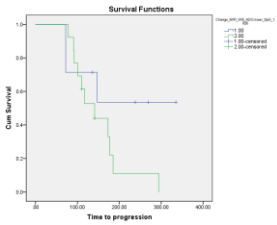
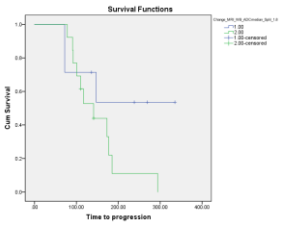
		%ΔADC$_{mean}^{WB}$	%ΔADC$_{median}^{WB}$
KM plot			
Dichotomisation		+1.626%	+1.8%
Median PFS in days (95% CI)	≤	0 (0-0)	0 (0-0)
	>	141 (91.7-190.3)	141 (91.7-190.3)
p-Value (Log Rank)		0.144	0.144

Table 13-24: WB DW-MRI; PFS Analysis with optimised dichotomisation; %Δ WB ADC Parameters; All Patients; KM Analysis Survival Table

KM analysis of the prostate cancer lesions only (Fig.13-20) demonstrates no significant separation of the survival curves, showing no utility in these patients for prediction of PFS with the %Δ ADC^{WB} parameters.

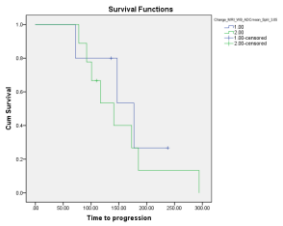
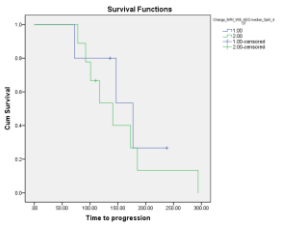
%ΔADC$_{meanWB}$ (prostate only)	%ΔADC$_{medianWB}$ (prostate only)
	
Dichotomised at +3.05% Low (1) - blue line High (2) - green line	Dichotomised at +4.07% Low (1) - blue line High (2) - green line
p=0.528	p=0.528

Figure 13-20: WB DW-MRI; PFS Analysis; %Δ WB ADC Parameters; Prostate patients only; KM Analysis

13.3.3.2 %ΔtDV$_{WB}$

	%ΔtDV$_{WB}$
Cox regression p-value [HR (95%CI)]	0.378 [0.995 (0.984-1.006)]

Table 13-25: WB DW-MRI; PFS Analysis; %Δ WB tDV; All Patients; Cox regression analysis

No significant statistical correlation has been identified using a Cox regression (Tbl.13-25) for these patients between the size of change of the whole-body tDV and their PFS. The data has been dichotomised at the median %ΔtDV$_{WB}$ has been used for KM PFS analysis (Tbl.13-26).

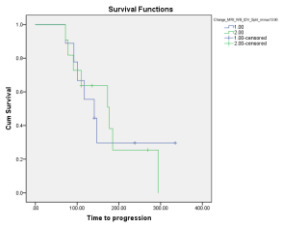
		%ΔtDV^{WB}
KM plot		
Dichotomisation		-13.9%
Median PFS in days (95% CI)	≤	141 (70.9-211.1)
	>	177 (92.2-261.8)
p-Value (Log Rank)		0.955

Table 13-26: WB DW-MRI; PFS Analysis; %Δ WB tDV; All Patients; KM Analysis Survival Table

No significant separation of the PFS curve has been identified for these patients using log rank analysis. Analysis of just the prostate cancer patients also demonstrates no significant separation of the survival curves (p=0.165).

13.3.3.3 %Δ Heterogeneity Parameters

	$\%\Delta ADC_{entropyWB}$	$\%\Delta ADC_{energyWB}$
Cox regression p-value **[HR (95%CI)]**	**0.705** [0.996 (0.975-1.018)]	**0.41** [1.001 (0.999-1.002)

Table 13-27: WB DW-MRI; PFS Analysis; %Δ WB ADC Heterogeneity Parameters; All Patients; Cox regression analysis

No significant correlation between the size of the %Δ whole-body heterogeneity parameters and the patients' PFS has been identified using Cox regression analysis (Tbl.13-27).

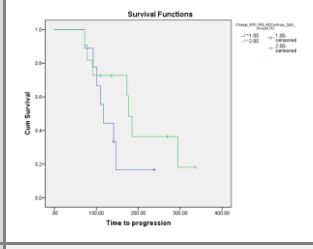
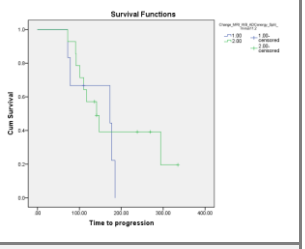
		$\%\Delta ADC_{entropy}^{WB}$	$\%\Delta ADC_{energy}^{WB}$
KM plot			
Dichotomisation		-0.71%	-11.2
Median PFS in days (95% CI)	≤	117 (96.5-137.5)	173 (0-359.2)
	>	177 (160.9-193.1)	141 (96.4-185.6)
p-Value (Log Rank)		0.186	0.460

Table 13-28: WB DW-MRI; PFS Analysis; %Δ WB ADC Heterogeneity Parameters; All Patients; KM Analysis Survival Table

There is no significant separation of the PFS curves either %Δ ADC heterogeneity parameter (Tbl.13-28). Analysis of only the prostate cancer patients (Tbl.13-29) has also failed to demonstrated PFS prognostic utility from these whole-body parameters.

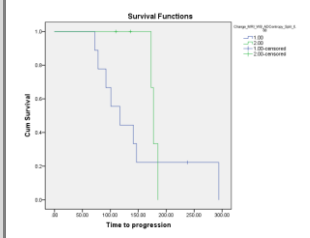
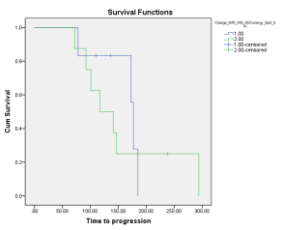
		$\%\Delta ADC_{entropy}^{WB}$ (Prostate only)	$\%\Delta ADC_{energy}^{WB}$ (Prostate only)
KM plot			
Dichotomisation		0.88	6.41
Median PFS in days (95% CI)	≤	117 (70.3-163.7)	177 (170.4-183.6)
	>	177 (170.6-183.4)	117 (61.6-172.4)
p-Value (Log Rank)		0.361	0.716

Table 13-29: WB DW-MRI; PFS Analysis; %Δ WB ADC Heterogeneity Parameters; Prostate patients only; KM Analysis Survival Table

13.4 Discussion

		ADC_{mean}^{WB}	ADC_{median}^{WB}	tDV^{WB}	$ADC_{entropy}^{WB}$	ADC_{energy}^{WB}
Baseline Scan	**Tumour**	X	X	X	X	X
	Response	X	X	X	X	X
	↑OS	↓ P pts - p=0.212	↓ P pts - p=0.064	↓ All pts – p=0.134 P Pts – p=0.164	X	X
	↑PFS	X	X	↓ P pts – p=0.036	↓ All pts – p=0.0002 P pts – p=0.017	↑ All pts – p=0.022 P pts – p=0.166
%Δ	**Tumour**	X	X	X	X	X
	Response	X	X	X	X	X
	↑OS	↓ All pts p=0.011 P pts – p=0.049	↓ All pts p=0.032 P pts – p=0.128	X	↑ Pts – p=0.087	X
	↑PFS	X	X	X	X	X

Table 13-30: WB DW-MRI: Summary Table

In this technical development study, successful whole-body quantification of DW-MRI has been achieved. The parameters used in this analysis have demonstrated their individual utility for bone metastases at the lesion level, but also at the patient level when target lesions have quantified within patients.

Target lesion selection is often arbitrary, and quantification of an isolated lesion does not necessarily take into account the inter-lesion heterogeneity known to exist in patients with bone metastases. Whole-body quantification gives an opportunity to consider tumour burden and the impact of inter-lesion heterogeneity.

The per-lesion analysis identified the ADC of individual breast and prostate cancer bone metastases were statistically different, with a lower ADC identified with prostate cancer metastases. However, no such difference has been identified with the whole-body analysis. This may represent the known heterogeneity of bone metastases within individual patients; within all patients there will be bone metastases with relatively high and low ADC parameters and the whole-body parameter may be diluted by this heterogeneity. The known differences between the breast and prostate cancer metastases may also be hidden by noise in the whole-body parameter. Within the whole-body segmentation there are probably voxels not associated with bone metastases. No utility for predicting a treatment response has been shown for these patients from the baseline ADC WB parameters.

The per-lesion analysis for these patients demonstrated a predictive utility from the %Δ ADC parameters; prostate cancer bone metastases with a more positive change in ADC between the scans more likely to be in patients with a treatment response. However, the whole-body parameters have not demonstrated an association with the 24-week treatment response. The per-lesion analysis also clearly demonstrated the heterogeneous response of individual bone metastases within individual patients. Whole-body quantification was embarked upon as a quantification method incorporating this heterogeneity, but perhaps correlating the more dominant change in ADC within the skeleton with the whole-body ADC parameters with the patient's overall clinical outcome. This has not been demonstrated with the whole-body ADC parameters. This may be due to inter-lesion heterogeneity or and image noise diluting the predictive data.

However, the whole-body ADC parameters have a trend towards significance for predicting overall-survival; in this group of patients, those with a lower baseline ADC_{median}^{WB} had a prolonged OS. This association was also found with per-lesion analysis. The lower ADC of the prostate cancer lesions suggest ADC is lower with sclerosis[234]. This OS advantage may suggest the degree of sclerosis in these prostate cancer patients is a predictor of OS. This has been shown in breast cancer patients previously[235]. A decrease in ADC after treatment was associated with an OS benefit with per-lesion analysis, perhaps reflecting further sclerosis associated with healing bone metastases. This has been also demonstrated significantly with the whole-body ADC parameters; patients with a decrease in ADC_{mean}^{WB} between the scans have an OS benefit (p=0.011).

Although the ADC^{WB} parameters have predicted for OS, there is no evidence to demonstrate they predict PFS for these patients, either from the baseline imaging or percentage change between the scans. The per-lesion analysis also failed to identify an association from the baseline imaging, but there was a trend towards prediction of PFS. Prostate cancer patients with decrease in ADC tended towards a PFS benefit; this has not been identified with whole-body analysis.

The whole-body volume of bone metastatic burden has failed to show prognostic utility for these patients. The per-lesion analysis did not show a statistically significant association between the baseline tDV and 24-week treatment response, but the largest lesions were in patients who failed to have a sustained treatment response. Whole-body quantification was expected to amplify this measure, resulting in a clinically useful parameter. However, this has not been demonstrated. Similarly, no association between the change in whole-body tDV and the 24-week treatment response for these patients.

The baseline tDV^{WB} shows a trend for prognosis in the prostate cancer lesions; lower skeletal burden of disease predicts an OS and PFS benefit. No predictive utility for OS or PFS has been identified by the change in tDV between the scans. Biologically a prolonged OS might be expected in patients with a lower skeletal burden of disease; Perez-Lopez et al showed, for prostate cancer patients, a lower baseline tDV was associated with an OS benefit[240]. Other imaging modalities have shown the disadvantageous effect of bone metastasis burden on OS[222,223,242]. The prediction of PFS is difficult to explain; the tDV parameters have not shown utility for predicting the overall clinical response to treatment, but perhaps a lower baseline volume of disease is associated with a less aggressive behaviour of tumour resulting in a prolonged time to clinically confirmed progression.

The baseline or %Δ ADC^{WB} heterogeneity parameters in these patients did not predict the treatment response. The per-lesion analysis also did not show utility.

The %Δ ADC heterogeneity parameters however seem to show predictive utility for anticipating the OS; patients with an increase in $ADC_{entropy}^{WB}$ or fall in ADC_{energy}^{WB} have an OS benefit (p=0.017 and 0.067 respectively). This statistical significance is not maintained when only the prostate cancer patients are analysed. For the prostate cancer analysis, there is a trend towards an OS benefit with %Δ $ADC_{entropy}^{WB}$ (p=0.087). There was no demonstrable difference in the %Δ ADC heterogeneity parameters between the tumour groups identified for these patients, and the

tumour subgroup analysis limits statistical comparisons by decreasing the sample size. The patients with a change in the whole-body heterogeneity parameters towards heterogeneity have an OS advantage. A fall in whole-body ADC is associated with an OS benefit, and it might be the same underlying processes are influencing the quantification parameters; sclerosis may be causing increased image restriction, and resulting in a more heterogeneous voxel distribution within the VOI.

PFS for these patients has been predicted with the baseline whole-body ADC heterogeneity parameters, showing a 193-day median PFS benefit for the patients with more homogeneous whole-body parameters (lower $ADC_{entropy}^{WB}$ (p=0.0002) and higher ADC_{energy}^{WB} (70-day PFS benefit p=0.022)). This association was maintained for the prostate cancer patients, although with fewer patients at sub-group analysis the statistical associations are not as strong. However, the change in ADC heterogeneity parameters between the scans have not demonstrated PFS predictive utility.

This study has demonstrates successful segmentation of only bone metastases from the whole imaged skeleton for quantification. The inter-lesion heterogeneity of bone metastases is clear, and has been demonstrated with the per-lesion analysis of these patients. Functional imaging modalities give the opportunity to interrogate the metabolic characteristics of tumours and the peri-tumour environment, and perhaps offer assessment of treatment responses of tumours before morphological changes take place. Quantification of DW-MRI images can identify predictive and prognostic behaviours of malignancies. Quantification in the clinical environment requires selection of a target lesion (or a few target lesions) to act as a surrogate marker. Which lesions are the most representative of the clinically most relevant disease is unknown. Whole-body quantification might eliminate the necessity for arbitrary selection of lesions, allowing quantification of all bone lesions.

This study has not, however, shown as much predictive and prognostic utility of the whole-body parameters as was suggested with the per-lesion analysis or per-patient target lesion analysis, and therefore with this methodology and these parameters the whole-body quantification does not show utility. However, the parameters do show trends towards those seen with the per-lesion and per-patient analyses. The prognostic discrimination has been diluted from the per-lesion analysis and perhaps improvement of the methodology could improve the clinical information gained from whole-body quantification.

There is likely to be significant noise in the image, with voxels included not represent bone metastases. The segmentation methodology was based on a small study using a high b-value image (b value = 1400) to auto-segment areas of high signal within the scan. This segmentation technique was based on per-lesion analysis. The method used for this study then required manual slice-by-slice deletion of non-bone voxels. This process was too time consuming to have clinical utility, and is likely to have introduced error. Identification of all bone metastatic disease was the aim, but there is no gold standard against which to compare the methodology. For this study the individual VOI parameters were weighted based on the volume of the VOI; other

methods might be to exclude VOIs below a certain threshold to try and reduce the impact of image noise.

Blackledge et al. have developed a semi-automatic segmentation method using statistical methods to segment bone metastases from whole-body diffusion weighted imaging, aiming to provide global measurement of tDV and ADC, specifically examining the bone metastases of breast and prostate cancer. Their initial feasibility study as published in 2014, after the design of the study reported in this thesis. They reported on 11 patients with breast or prostate cancer bone metastases, and identified an increase in global ADC (gADC) after treatment compared to non-responding patients, and non-responding patients demonstrated an increase in tDV[154]. None of the parameters in this study demonstrate an association with the 24-week treatment response. The Blackledge study had a much greater time between MRI images than was designed into this thesis study with an average of 22 weeks (range 10.5 - 38weeks). The time for response assessment is not reported in their paper. The scans at 8-12 weeks might be too early to demonstrate clinically relevant changes in signal.

In 2016 the same group published inter- and intra-observer repeatability of their semi-automatic segmentation methodology, concluding excellent repeatability for gADC and gtDV. I have not assessed the reproducibility and repeatability of the technique used in this study; this would be more relevant had more clinically relevant utility been identified.

A limiting factor in this analysis has been the small sample size; small differences in the parameters may have clinical relevance, but the recruitment thus far is underpowered to demonstrate such differences. The studies are still open for recruitment, and there are now more patients who could be included for future analysis. It would be clinically relevant to correlate the parameter changes with PSA, markers of bone turn over (e.g. ALP), and it would be relevant to quantify the sclerotic nature of the metastases (e.g. recording the Hounsfield Units of the CT component of the scans).

The underlying physical and morphological characteristics resulting in the parameter changes are unclear. Histological comparisons of the ADC parameters with prostate cancers have shown correlation with Gleason grade and more aggressive histology[220]. However, the image signal from bone metastases is complicated by the associated bone stroma and marrow, both of which will influence the DW-MRI signal, and will be affected by the treatment, and by the mode of treatment. The per-lesion analysis demonstrated significant differences between the characteristics of the parameters from prostate and breast bone metastases. It is unclear if the differences are down to the tumour interaction with bone, or if the cytotoxic therapy has a different impact on the signal from the bone marrow compared to hormonal therapies.

Chapter 14 Discussion

Bone metastases are common sequelae of malignancy, and cause patients significant morbidity. Effective treatments are available, and more are being developed. Whereas soft tissue metastatic deposits can be assessed by changes in size, assessing treatment responses in bone metastases in an accurate and timely manner is much more complicated because morphological responses to therapy are slow and inconsistent. Bone metastases are commonly excluded from inclusion as index lesions in clinical studies; patients with metastatic disease confined to the bone are therefore poorly represented by clinical trials.

The three prospective clinical trials, FAB-P, FAB-IE and FAB-B have been designed to interrogate modern functional imaging techniques and the potential predictive and prognostic data quantified from the images.

This thesis has reported the initial patients recruited to these studies, focusing on ^{18}F-Fluoride PET imaging and DW-MRI, with the primary aims of demonstrating the predictive and prognostic utility from these images, identifying clinically applicable quantification parameters, and developing quantification methods which can be used for the final data analyses of the FAB studies.

What is evident from all the analyses is the differences in imaging characteristics between the breast and prostate cancer pathologies. The pre-treatment functional imaging characteristics are different between these diagnoses, and how the quantification parameters change with treatment are significantly different. Future studies should be recommended to focus on individual pathologies, or ensure a sufficient population to enable multivariate analysis (which has not been possible with this analysis). Prostate bone metastases are primarily osteoblastic. Breast cancer lesions usually cause a mixed osteoblastic/osteolytic pattern of disease; the underlying biology contributing to these phenotypes is likely to be the largest contributing factor to the differences identified.

Intra-patient inter-lesion heterogeneity is noted throughout these results, both at the baseline scans, and also heterogeneity of responses to therapy. Within one patient lesions can be identified with diametrically opposed changes of quantification parameters with treatment. If a single lesion were to be used as the clinical index marker quite different conclusions could be drawn depending on the selection of the lesion. Which lesion is the most clinically relevant is unclear. For both the ^{18}F-Fluoride PET and DW-MRI, clear associations between the quantification parameters and clinical outcomes were made with per-lesion analysis, but much of this utility was not identified at the per-patient level. This may reflect the much smaller study sample of the per-patient analysis affecting the size of difference identifiable with statistical comparisons, but the inter-lesion heterogeneity may also have a larger impact when per-patient or whole-body analysis is used. However, it is this level of analysis required for clinical application. Better approaches to per-patient analysis may be necessary to extract the clinical predictive and prognostic signal identified with per-lesion analysis.

Discussion

The ^{18}F-Fluoride PET per-lesion analysis provided 62 metastases for analysis, with 24 breast cancer lesions and 38 prostate cancer lesions. Only 2 of the prostate cancer lesions represented a patient with a treatment response, which has limited some of the tumour-subgroup analyses. No baseline ^{18}F-Fluoride parameter has been shown in this analysis to predict the 24-week treatment response. However, an increase in SUV_{max} (p=0.048) (trend with SUV_{mean} (p=0.065)) after 8 weeks of treatment was predictive of the 24-week treatment response for the breast cancer lesions. The change in ^{18}F-Fluoride metabolic volume did not predict treatment response, and for some patients with a treatment response the volume of tracer uptake increased, which was unexpected and may reflect a healing osteoblastic driven response within the peri-tumour bone. For the breast cancer lesions, there was a trend towards significance for using the heterogeneity parameters to identify a treatment response at 8 weeks, with a response more likely in those with a change towards homogeneity (decrease in $SUV_{entropy}$ (p=0.140) and increase in SUV_{energy} (p=0.093). Ki is not predictive. There is significant correlation between %Δ SUV parameters and heterogeneity parameters, but there are too few cases for multivariate analysis to identify if they perform as independent predictors.

The baseline ^{18}F-Fluoride scan, with per-lesion analysis, showed there is prognostic information to be elicited. Lesions with lower $SUV_{mean/max/peak}$ (p=0.0002), lower $SUV_{entropy}$ (p=0.0002) and higher SUV_{energy} (p=0.0004) (suggesting a less random and more uniform distribution of SUV voxels), and a lower Ki (p=0.001) were associated with a longer OS. There is significant and strong correlation between these parameters, but multivariate analysis to identify independent prognostic factors is not feasible with this study size. This effect seemed independent of tumour subgroup analysis. These parameters were not shown to be prognostic of PFS.

How these ^{18}F-Fluoride parameters changed with treatment was also associated with OS prognostication. Lesions with an increase in $SUV_{mean/median/peak}$ (p=0.016,0.001,0.014), increase in $SUV_{entropy}$ (p=0.014) and fall in SUV_{energy} (p=0.008) were associated with an OS benefit; these associations were also identified when only the prostate lesions were analysed. As previously mentioned, these parameters are all closely correlated, and further analysis of a larger study group would be necessary to identify independence of these prognosticators. Volumetric parametersi demonstrated no utility for these patients. An increased ^{18}F-Fluoride avidity demonstrates increased osteoblastic activity, and the association with OS might suggest healing of the metastatic deposits, although there was no association for these parameters with the 24-week response assessment or PFS. This may reflect a different biology of the underlying malignancies, with a longer course of the disease despite no demonstrable benefit in PFS following this line of therapy.

None of this predictive and prognostic utility was carried through to the per-patient analysis. There were too few patients to perform tumour-subgroup analyses. Differences were suggested between the response groups, a larger study group might confirm these differences. Whether these differences were due the tumour group or the biology of the tumour deposits, the per-patient analysis measured these differences. If there is clinically predictive or prognostic utility a larger group might identify this. It is not clear if it is the target lesion selection, inter-lesion heterogeneity,

or the reduction in sample size preventing the predictive and prognostic utility shown at per-lesion analysis being continued through to analysis at the patient level.

The per-patient analysis of the ^{18}F-Fluoride PET did however provide an opportunity to demonstrate the significant impact the VOI definition method has on the resulting parameters. As expected, SUV_{max} and SUV_{peak} are much less affected by the choice of VOI definition method than the other parameters. SUV_{mean}, with the potential benefit of representing the full VOI voxel population (rather than just a sample, such as SUV_{max} or SUV_{peak}) is significantly affected by the choice of iso-contour. The statistical FLAB method for VOI definition has produced parameters most similar to the 27% iso-contour, used as the method felt most likely to represent the PET volume matching the sclerotic component on CT imaging.

The %Δ ^{18}F-Fluoride PET parameters show less variation across the VOI methods than the baseline parameters for the relative iso-contour methods based on the SUV_{max} of each lesion. The parameters measured using the same SUV iso-contour applied to the baseline imaging showed much greater differences.

It had been an aim to identify the VOI method showing the best clinical utility; unfortunately, due to the lack of predictive or prognostic associations identified with per-patient analysis, it has not been possible to select the better methodology. However, consistency of methodology is clearly a priority. The higher degree of consistency of the %Δ parameters between relative iso-contour VOI methods suggests these parameters might be more comparable between studies irrespective of the chosen iso-contour for VOI definition, providing the same method is used for the baseline and subsequent scan for calculating the percentage change. The FLAB method most similar to the 27% iso-contour is shown to best match sclerosis. Others have shown FLAB to be more reproducible than iso-contour methods - a clear advantage. Current computation limitations restrict the use of FLAB for whole-body analysis.

There is limited evidence against which to compare the findings of the ^{18}F-Fluoride analysis. Changes in ^{18}F-Fluoride SUV of prostate bone metastases correlates with PSA[140,195], and deceases in breast cancer metastasis SUV following successful treatment[196]. Others have shown no predictive utility[191], nor prognostic use [204,206]. The skeletal burden of disease can predict OS[204].

The baseline per-lesion analysis of the DW-MRI scans has suggested more predictive and prognostic potential than the ^{18}F-Fluoride scans, but this analysis had the benefit of a larger study population. The ADC parameters (mean and median) of the baseline imaging were significantly different between the breast and prostate tumours. This may represent the increased sclerotic component expected in osteoblastic prostate cancer metastases. The prostate lesions on average had a lower baseline ADC, perhaps with the sclerotic reaction causing diffusion restriction. These baseline ADC parameters have not been shown to predict treatment response, but the heterogeneity parameters showed a trend towards predictive utility. Breast lesions with a lower $ADC_{entropy}$ ($p=0.069$) and higher ADC_{energy} ($p=0.059$) (i.e. more homogeneity of SUV voxel distribution) were more likely to be identified in patients with a treatment response.

An increase in ADC following successful treatment has predicted a treatment response for the prostate cancer lesions ($p=0.029$). A similar trend was suggested for the breast cancer lesions,

but there were too few to achieve statistical significance. Others have reported an increase in ADC following successful therapy in skeletal malignancies [58,148,157,229-231], and specifically prostate cancer bone metastases[151-153,232,233]. Volumetric parameters have shown no predictive utility from per-lesion analysis of the DW-MRI data.

The DW-MRI data, at the per-lesion analysis level, has also shown great promise for clinical prognostic application. A lower baseline ADC predicts for an OS benefit of the prostate cancer lesions ($p \leq 0.05$), but did not predict PFS in these patients. ADC_{energy} was also shown to have utility, with more uniform values (higher) predicting an OS benefit for the prostate cancer lesions and PFS for breast cancer lesions.

An increase in ADC predicted a clinical treatment response for the prostate cancer lesions, but in the same lesions a fall in ADC between scans is associated with an OS benefit ($p=0.005$). An association between treatment response and survival might be anticipated, but this result does not suggest this for these patients. Without fully understanding what is causing the ADC signal change in the bone metastases following treatment it is difficult to explain this finding. There was a trend seen in this analysis for a PFS advantage for prostate cancer lesions showing a fall in ADC, but for breast cancer lesions with an increase in ADC. Similarly, a fall in $ADC_{entropy}$ and rise in ADC_{energy} between scans predicts a PFS benefit for the prostate cancer lesions, but the converse pattern is associated with a PFS benefit for the breast cancer lesions. The underlying pathophysiology of the tumour, the tumour-bone reaction and the treatment is therefore influencing the ADC signal, and further work will be necessary to understand the pathophysiological processes behind the signal changes.

Despite the clinical utility suggested from the per-lesion analysis, the DW-MRI has proven less useful at the per-patient analysis. This may be due to the much smaller study size, or may reflect the target-lesion selection is not representing the clinically relevant disease sufficiently. The inter-lesion heterogeneity may be diluting the predictive and prognostic signal when up to 5 lesions are quantified per patient. The only parameter to suggest utility was the tDV, where an increase in tDV predicted a treatment response for the prostate cancer patients ($p=0.026$). An increase in size suggests the tDV is not a measure of the tumour cells, but is also measuring the associated bone microenvironment, and changes in the bone following treatment. A healing osteoblastic reaction causing sclerosis could lead to image restriction, and a larger VOI defined on the extrapolated b-value image.

There is no evidence in the literature to suggest ADC measurements predict the ultimate clinical treatment response of bone metastases. There is, however, an increasing body of evidence to suggest a lower ADC in tumour tissue (not bone metastases) before therapy is associated with an increased response rate[145,213-219]. There is evidence to suggest the volume of osseous disease is prognostic in prostate cancer[222,223], but there is no evidence showing the baseline volume of bone metastases is predictive of treatment response. A correlation between heterogeneity and the prostate cancer Gleason grade has been reported[172,176,177,224], and Gleason grading is predictive of the clinical course of the disease. Others have reported a rise in ADC following a

treatment response in malignancies involving the skeleton[58,148,157,229-231], and specifically prostate cancer bone metastases[151-153,232,233].

The whole-body analyses were an attempt to identify methods for segmenting out all the skeletal disease; hypothesising clinical utility for quantification of the majority of skeletal disease because identification of a single, or few, clinically relevant target lesions is difficult and uncertain.

The skeletal segmentation of the PET scan from the CT component of the PET/CT is perhaps more clinically useful, although with the ^{18}F-Fluoride tracer this technique did not yield significant predictive or prognostic information. This technique was relatively quick and easy to apply, and might be useful for other PET tracers where there is more non-skeletal physiologic uptake, or for metastatic disease with bone and soft tissue disease. Prognostic information was gained from the whole-body ^{18}F-Fluoride PET analysis not seen with the per-patent analysis. An increase in SUV_{meanWB} between scans suggested a PFS benefit for the prostate cancer patients (p=0.025), as did a fall in TLA (p=0.025). The MTV showed a trend towards benefit, but the combination of MTV and SUV_{meanWB} has resulted, for these patients, in a more discriminatory parameter in TLA.

The whole-body DW-MRI analysis has shown a PFS benefit for patients with a lower baseline tDV_{WB}, $ADC_{entropyWB}$ and higher $ADC_{energyWB}$ (i.e. smaller tumour burden and more homogeneous ADC voxel distribution). A fall in ADC_{meanWB} and $ADC_{medianWB}$ between scans is predictive of an OS benefit, as identified at the per-patient analysis.

The inter-lesion heterogeneity seen within patients is a justification for interrogating the whole-skeletal burden. There are many other ways this data could be analysed, including descriptions of the whole skeletal population of disease, such as kurtosis and skewness. These could be investigated in future analyses.

However, this thesis has shown whole skeletal segmentation of functional imaging data is readily achievable, and this may have significant clinical relevance. One more immediate next analysis will be to examine the volume of the segmented CT component of the PET/CT scans, with the aim of concluding the percentage skeletal burden of metastases as other groups have clearly demonstrated the prognostic value of this approach[205].

There are limitation with these studies. The study population, particularly at the per-patient analysis, is relatively small, particularly as the influence of the tumour sub-groups is significant. The same patients, and lesions, have been used for all these studies; correction could be applied to the statistical level of significance to take these multiple analyses into account, but this has not been applied as the studies have been exploratory in nature. This level of scrutiny will however be necessary when the trials reach full recruitment for final analysis.

When designing imaging methodology the repeatability and reproducibility of approaches is a vital component; before selecting the VOI methodology for final data analysis it will be necessary to consider both of these. Although the VOI methods have not shown superiority from a clinical perspective in these preliminary studies, there are significant differences between the resulting parameters; repeatability and reproducibility might be favourable with a particular VOI methodology.

Discussion

Future developments of this analysis will be to include correlation with tumour markers and markers of bone turnover; particularly with the differences identified between the tumour types it will be necessary to identify changes in bone turnover resulting in the signal changes identified. Sclerosis is probably impacting the ADC signal (and, by the tracer mechanism, ^{18}F-Fluoride uptake); comparison of the ADC with a measure of the lesion sclerosis (for example, using Hounsfield unit measurements from the CT scan) may indicate further the bone changes leading to the signal variations. Histological correlation would be the gold-standard in this setting.

Chapter 15 Conclusions

Quantification of ^{18}F-Fluoride PET and DW-MRI images of bone metastases, it was hypothesised, would enable earlier prediction of the 24-week response assessment. This initial analysis certainly demonstrates the ability of both imaging techniques to discern a treatment response at only 8 to 12 weeks after starting therapy, implying clinical utility as imaging biomarkers of early response. In addition, prognostic utility has been suggested for both OS and PFS.

Breast and prostate cancer bone metastases have demonstrably different imaging characteristics; the baseline parameters were different, and how parameters changed with treatment was dependent on the tumour-specific pathophysiology of the bone lesions. This was seen with the changes in ADC; a fall in ADC was associated with a PFS benefit for prostate cancer lesions, but the converse was seen with breast cancer lesions. The specific cellular processes affecting the ADC parameters are uncertain, but clearly tumour-specific analyses of the scans need to be developed, and a unified approach for all bone metastases is misguided.

The impact on the ^{18}F-Fluoride parameters by the VOI delineation methods is significant. The parameters describing the full VOI voxel population (SUV_{mean}, volumetric and heterogeneity parameters) are affected more than those describing only a sample (SUV_{max} and $SUV_{peak)}$, showing a clear practical advantage for these parameters. It has not been possible to identify a superior VOI method. The per-patient study population was perhaps too small to identify the predictive and prognostic utility suggested from per-lesion analysis. Consistency of methodology is vital for meaningful comparisons to be made.

Intra-patient inter-lesion heterogeneity has been a predominant feature through this analysis. Lesions within individual patients have been shown to have diametrically opposed changes in parameters following therapy; quantification using only target lesions is reliant on identification of the most clinically relevant lesions, and the right approach to this is unclear. The per-patient analysis did not show the same response and survival discrimination identified with per-lesion analysis. The most clinically indicative bone metastases may not have been selected for analysis, or the smaller sample size may be limiting the analysis. Inclusion of all skeletal bone metastases for whole-body quantification was hypothesised to address this target lesion selection problem.

Whole-body quantification methods, producing parameters describing the whole-skeletal metastatic burden, have been shown to be readily achievable. Whole-body ^{18}F-Fluoride PET analysis identified discriminatory image signal, showing expected differences between the prostate and breast cancer lesions. However, clinical utility was not shown, perhaps limited by the sample size and uncertainty of the best segmentation threshold to minimise non-malignant disease in the quantification.

Whole-body DW-MRI imaging, with a larger study sample, demonstrated prognostic utility, both of overall and progression free survival, independent of the tumour groups. This was not with greater accuracy than target lesion per-patient analysis.

This thesis was expected to guide methodological approaches for the final data analysis of the FAB trials. The parameters selected have all suggested varying degrees of utility, but describe

the voxel distribution in a variety of ways maximising the likelihood of extracting clinically relevant data from the scans. The choice of VOI definition method for ^{18}F-Fluoride quantification is fundamental, and further work will be necessary to identify the best method to take forward. Reproducibility and repeatability will inform this decision. FLAB, with superior reproducibility to iso-contour methods, indicates an advantage.

The next key step is to confirm quantification methodologies that translate the predictive and prognostic utility of these functional imaging methods, demonstrated at the per-lesion level, into applicable methods for prediction and prognostication of cancer patients undergoing treatments in the clinic.

Chapter 16 References

1. Coleman RE. Metastatic bone disease: clinical features, pathophysiology and treatment strategies. *Cancer treatment reviews* 2001; **27**(3): 165-76.
2. Coleman RE, Rubens RD. Bone metastases and breast cancer. *Cancer treatment reviews* 1985; **12**(4): 251-70.
3. Hofbauer LC, Rachner T, Singh SK. Fatal attraction: why breast cancer cells home to bone. *Breast cancer research : BCR* 2008; **10**(1): 101.
4. Paget S. The distribution of secondary growths in cancer of the breast. *Lancet* 1889; **133**(3421): 571-3.
5. Krasnow AZ, Hellman RS, Timins ME, Collier BD, Anderson T, Isitman AT. Diagnostic bone scanning in oncology. *Semin Nucl Med* 1997; **27**(2): 107-41.
6. Jacobson AF, Fogelman I. Bone scanning in clinical oncology: does it have a future? *Eur J Nucl Med* 1998; **25**: 1219-23.
7. Plunkett TA, Rubens RD. The biology and management of bone metastases. *Critical reviews in oncology/hematology* 1999; **31**(1): 89-96.
8. Mundy GR. Malignancy and the skeleton. *Hormone and metabolic research = Hormon- und Stoffwechselforschung = Hormones et metabolisme* 1997; **29**(3): 120-7.
9. Sterling JA, Edwards JR, Martin TJ, Mundy GR. Advances in the biology of bone metastasis: how the skeleton affects tumor behavior. *Bone* 2011; **48**(1): 6-15.
10. Dougall WC. Molecular pathways: osteoclast-dependent and osteoclast-independent roles of the RANKL/RANK/OPG pathway in tumorigenesis and metastasis. *Clinical cancer research : an official journal of the American Association for Cancer Research* 2012; **18**(2): 326-35.
11. Ghanem N, Uhl M, Brink I, et al. Diagnostic value of MRI in comparison to scintigraphy, PET, MS-CT and PET/CT for the detection of metastases of bone. *Eur J Radiol* 2005; **55**(1): 41-55.
12. Statistics OfN. Cancer Incidence and Mortality in the United Kingdom, 2008-10. 2012. http://www.ons.gov.uk/ons/dcp171778_289890.pdf (accessed 17/04/13.
13. Siegel R, Ma J, Zou Z, Jemal A. Cancer statistics, 2014. *CA: a cancer journal for clinicians* 2014; **64**(1): 9-29.
14. DeVita VT, Hellman S, Rosenberg SA, editors. Cancer: Principles & Practice of Oncology. Philadelphia, USA: Lippincott Williams & Wilkins; 2005.
15. Bubendorf L, Schöpfer A, Wagner U, et al. Metastatic patterns of prostate cancer: An autopsy study of 1,589 patients. *Human Pathology* 2000; **31**(5): 578-83.
16. Nemeth JA, Harb JF, Barroso U, He ZQ, Grignon DJ, Cher ML. Severe combined immunodeficient-hu model of human prostate cancer metastasis to human bone. *Cancer Research* 1999; **59**(8): 1987-93.
17. Lehr JE, Pienta KJ. Preferential adhesion of prostate cancer cells to a human bone marrow endothelial cell line. *Journal of the National Cancer Institute* 1998; **90**(2): 118-23.
18. Morrissey C, Vessella RL. The role of tumor microenvironment in prostate cancer bone metastasis. *Journal of cellular biochemistry* 2007; **101**(4): 873-86.
19. Koeneman KS, Yeung F, Chung LWK. Osteomimetic properties of prostate cancer cells: A hypothesis supporting the predilection of prostate cancer metastasis and growth in the bone environment. *Prostate* 1999; **39**(4): 246-61.
20. Saad F. Zoledronic acid significantly reduces pathologic fractures in patients with advanced-stage prostate cancer metastatic to bone. *Clinical prostate cancer* 2002; **1**(3): 145-52.
21. Garnero P, Buchs N, Zekri J, Rizzoli R, Coleman RE, Delmas PD. Markers of bone turnover for the management of patients with bone metastases from prostate cancer. *British journal of cancer* 2000; **82**(4): 858-64.
22. Heidenreich A, Hofmann R, Engelmann UH. The use of bisphosphonate for the palliative treatment of painful bone metastasis due to hormone refractory prostate cancer. *Journal of Urology* 2001; **165**(1): 136-40.

23. Statistics OfN. Cancer Registration Statistics, England, 2011. *Cancer Registration Statistics, England*; **26 June 2013**.
24. Kuchuk I, Hutton B, Moretto P, Ng T, Addison CL, Clemons M. Incidence, consequences and treatment of bone metastases in breast cancer patients—Experience from a single cancer centre. *Journal of Bone Oncology* 2013; **2**(4): 137-44.
25. Schröder J, Fietz T, Köhler A, et al. Treatment and pattern of bone metastases in 1094 patients with advanced breast cancer – Results from the prospective German Tumour Registry Breast Cancer cohort study. *European Journal of Cancer* 2017; **79**: 139-48.
26. Guise TA. Molecular mechanisms of osteolytic bone metastases. *Cancer* 2000; **88**(12 Suppl): 2892-8.
27. Kozlow W, Guise TA. Breast cancer metastasis to bone: mechanisms of osteolysis and implications for therapy. *Journal of mammary gland biology and neoplasia* 2005; **10**(2): 169-80.
28. Aapro MS, Coleman RE. Bone health management in patients with breast cancer: current standards and emerging strategies. *Breast (Edinburgh, Scotland)* 2012; **21**(1): 8-19.
29. Cardoso F, Harbeck N, Fallowfield L, Kyriakides S, Senkus E. Locally recurrent or metastatic breast cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. *Annals of oncology : official journal of the European Society for Medical Oncology / ESMO* 2012; **23 Suppl 7**: vii11-9.
30. Clemons M, Gelmon KA, Pritchard KI, Paterson AH. Bone-targeted agents and skeletal-related events in breast cancer patients with bone metastases: the state of the art. *Current oncology (Toronto, Ont)* 2012; **19**(5): 259-68.
31. Roelofs AJ, Thompson K, Ebetino FH, Rogers MJ, Coxon FP. Bisphosphonates: molecular mechanisms of action and effects on bone cells, monocytes and macrophages. *Current pharmaceutical design* 2010; **16**(27): 2950-60.
32. Fizazi K, Carducci M, Smith M, et al. Denosumab versus zoledronic acid for treatment of bone metastases in men with castration-resistant prostate cancer: a randomised, double-blind study. *Lancet* 2011; **377**(9768): 813-22.
33. Martin M, Bell R, Bourgeois H, et al. Bone-related complications and quality of life in advanced breast cancer: results from a randomized phase III trial of denosumab versus zoledronic acid. *Clinical cancer research : an official journal of the American Association for Cancer Research* 2012; **18**(17): 4841-9.
34. Hamaoka T, Madewell JE, Podoloff DA, Hortobagyi GN, Ueno NT. Bone imaging in metastatic breast cancer. *J Clin Oncol* 2004; **22**(14): 2942-53.
35. Roberts CC, Daffner RH, Weissman BN, et al. ACR appropriateness criteria on metastatic bone disease. *Journal of the American College of Radiology : JACR* 2010; **7**(6): 400-9.
36. Cuccurullo V, Cascini GL, Tamburrini O, Rotondo A, Mansi L. Bone metastases radiopharmaceuticals: an overview. *Curr Radiopharm* 2013; **6**(1): 41-7.
37. Choi J, Raghavan M. Diagnostic imaging and image-guided therapy of skeletal metastases. *Cancer control : journal of the Moffitt Cancer Center* 2012; **19**(2): 102-12.
38. O'Sullivan GJ, Carty FL, Cronin CG. Imaging of bone metastasis: An update. *World J Radiol* 2015; **7**(8): 202-11.
39. Bauerle T, Semmler W. Imaging response to systemic therapy for bone metastases. *Eur Radiol* 2009; **19**(10): 2495-507.
40. Norris DG. The effects of microscopic tissue parameters on the diffusion weighted magnetic resonance imaging experiment. *NMR in biomedicine* 2001; **14**(2): 77-93.
41. Padhani AR, Liu G, Koh DM, et al. Diffusion-weighted magnetic resonance imaging as a cancer biomarker: consensus and recommendations. *Neoplasia* 2009; **11**(2): 102-25.
42. Charles-Edwards EM, deSouza NM. Diffusion-weighted magnetic resonance imaging and its application to cancer. *Cancer imaging : the official publication of the International Cancer Imaging Society* 2006; **6**: 135-43.

43. Padhani AR. Diffusion magnetic resonance imaging in cancer patient management. *Semin Radiat Oncol* 2011; **21**(2): 119-40.
44. Vande Berg BC, Lecouvet FE, Michaux L, Ferrant A, Maldague B, Malghem J. Magnetic resonance imaging of the bone marrow in hematological malignancies. *Eur Radiol* 1998; **8**(8): 1335-44.
45. Sohaib SA, Cook G, Allen SD, Hughes M, Eisen T, Gore M. Comparison of whole-body MRI and bone scintigraphy in the detection of bone metastases in renal cancer. *Br J Radiol* 2009; **82**(980): 632-9.
46. Schmidt GP, Schoenberg SO, Schmid R, et al. Screening for bone metastases: whole-body MRI using a 32-channel system versus dual-modality PET-CT. *Eur Radiol* 2007; **17**(4): 939-49.
47. Pfannenberg C, Aschoff P, Schanz S, et al. Prospective comparison of 18F-fluorodeoxyglucose positron emission tomography/computed tomography and whole-body magnetic resonance imaging in staging of advanced malignant melanoma. *Eur J Cancer* 2007; **43**(3): 557-64.
48. Lecouvet FE, Geukens D, Stainier A, et al. Magnetic resonance imaging of the axial skeleton for detecting bone metastases in patients with high-risk prostate cancer: Diagnostic and cost-effectiveness and comparison with current detection strategies. *JOURNAL OF CLINICAL ONCOLOGY* 2007; **25**(22): 3281-7.
49. Ghanem N, Altehoefer C, Hogerle S, et al. Comparative diagnostic value and therapeutic relevance of magnetic resonance imaging and bone marrow scintigraphy in patients with metastatic solid tumors of the axial skeleton. *Eur J Radiol* 2002; **43**(3): 256-61.
50. Altehoefer C, Ghanem N, Hogerle S, Moser E, Langer M. Comparative detectability of bone metastases and impact on therapy of magnetic resonance imaging and bone scintigraphy in patients with breast cancer. *Eur J Radiol* 2001; **40**(1): 16-23.
51. Engelhard K, Hollenbach HP, Wohlfart K, von Imhoff E, Fellner FA. Comparison of whole-body MRI with automatic moving table technique and bone scintigraphy for screening for bone metastases in patients with breast cancer. *Eur Radiol* 2004; **14**(1): 99-105.
52. Nakanishi K, Kobayashi M, Takahashi S, et al. Whole body MRI for detecting metastatic bone tumor: comparison with bone scintigrams. *Magnetic resonance in medical sciences : MRMS : an official journal of Japan Society of Magnetic Resonance in Medicine* 2005; **4**(1): 11-7.
53. Venkitaraman R, Cook GJ, Dearnaley DP, et al. Does magnetic resonance imaging of the spine have a role in the staging of prostate cancer? *Clin Oncol (R Coll Radiol)* 2009; **21**(1): 39-42. Epub 2008 Nov 6.
54. Wang J, Takashima S, Takayama F, et al. Head and neck lesions: characterization with diffusion-weighted echo-planar MR imaging. *Radiology* 2001; **220**(3): 621-30.
55. Guo Y, Cai YQ, Cai ZL, et al. Differentiation of clinically benign and malignant breast lesions using diffusion-weighted imaging. *Journal of magnetic resonance imaging : JMRI* 2002; **16**(2): 172-8.
56. Kuroki Y, Nasu K, Kuroki S, et al. Diffusion-weighted imaging of breast cancer with the sensitivity encoding technique: analysis of the apparent diffusion coefficient value. *Magnetic resonance in medical sciences : MRMS : an official journal of Japan Society of Magnetic Resonance in Medicine* 2004; **3**(2): 79-85.
57. Woodhams R, Matsunaga K, Iwabuchi K, et al. Diffusion-weighted imaging of malignant breast tumors: the usefulness of apparent diffusion coefficient (ADC) value and ADC map for the detection of malignant breast tumors and evaluation of cancer extension. *Journal of computer assisted tomography* 2005; **29**(5): 644-9.
58. Byun WM, Shin SO, Chang Y, Lee SJ, Finsterbusch J, Frahm J. Diffusion-weighted MR imaging of metastatic disease of the spine: assessment of response to therapy. *AJNR American journal of neuroradiology* 2002; **23**(6): 906-12.
59. Brenner AI, Koshy J, Morey J, Lin C, DiPoce J. The bone scan. *Semin Nucl Med* 2012; **42**(1): 11-26.

60. Vijayanathan S, Butt S, Gnanasegaran G, Groves AM. Advantages and limitations of imaging the musculoskeletal system by conventional radiological, radionuclide, and hybrid modalities. *Semin Nucl Med* 2009; **39**(6): 357-68.
61. Talbot J, Paycha F, Balogova S. Diagnosis of bone metastasis: recent comparative studies of imaging modalities. *The quarterly journal of nuclear medicine and molecular imaging : official publication of the Italian Association of Nuclear Medicine (AIMN) [and] the International Association of Radiopharmacology (IAR), [and] Section of the So* 2011; **55**(4): 374-410.
62. Schirrmeister H, Glatting G, Hetzel J, et al. Prospective Evaluation of the Clinical Value of Planar Bone Scans, SPECT, and 18F-Labeled NaF PET in Newly Diagnosed Lung Cancer. *J Nucl Med* 2001; **42**: 1800-4.
63. Gnanasegaran G, Cook G, Adamson K, Fogelman I. Patterns, Variants, Artifacts, and Pitfalls in Conventional Radionuclide Bone Imaging and SPECT/CT. *Seminars in Nuclear Medicine* 2009; **39**(6): 380-95.
64. Quilty PM, Kirk D, Bolger JJ, et al. A comparison of the palliative effects of strontium-89 and external beam radiotherapy in metastatic prostate cancer. *Radiotherapy and oncology : journal of the European Society for Therapeutic Radiology and Oncology* 1994; **31**(1): 33-40.
65. Han LJ, Au-Yong TK, Tong WC, Chu KS, Szeto LT, Wong CP. Comparison of bone single-photon emission tomography and planar imaging in the detection of vertebral metastases in patients with back pain. *Eur J Nucl Med* 1998; **25**(6): 635-8.
66. Schirrmeister H, Guhlmann A, Elsner K, et al. Sensitivity in detecting osseous lesions depending on anatomic localization: planar bone scintigraphy versus 18F PET. *J Nucl Med* 1999; **40**: 1623-9.
67. Beheshti M, Langsteger W, Fogelman I. Prostate Cancer: Role of SPECT and PET in Imaging Bone Metastases. *Seminars in Nuclear Medicine* 2009; **39**(6): 396-407.
68. Sarikaya I, Sarikaya A, Holder LE. The role of single photon emission computed tomography in bone imaging. *Semin Nucl Med* 2001; **31**(1): 3-16.
69. Costelloe CM, Rohren EM, Madewell JE, et al. Imaging bone metastases in breast cancer: techniques and recommendations for diagnosis. *The lancet oncology* 2009; **10**(6): 606-14.
70. Kosuda S, Kaji T, Yokoyama H, et al. Does Bone SPECT Actually Have Lower Sensitivity for Detecting Vertebral Metastasis Than MRI? *The Journal of Nuclear Medicine* 1995; **37**(6): 975-8.
71. Helyar V, Mohan HK, Barwick T, et al. The added value of multislice SPECT/CT in patients with equivocal bony metastasis from carcinoma of the prostate. *European journal of nuclear medicine and molecular imaging* 2010; **37**(4): 706-13.
72. Barrington SF, Qian W, Somer EJ, et al. Concordance between four European centres of PET reporting criteria designed for use in multicentre trials in Hodgkin lymphoma. *European journal of nuclear medicine and molecular imaging* 2010; **37**(10): 1824-33.
73. Gallamini A, Hutchings M, Rigacci L, et al. Early interim 2-[18F]fluoro-2-deoxy-D-glucose positron emission tomography is prognostically superior to international prognostic score in advanced-stage Hodgkin's lymphoma: a report from a joint Italian-Danish study. *J Clin Oncol* 2007; **25**(24): 3746-52.
74. Lin C, Itti E, Haioun C, et al. Early 18F-FDG PET for prediction of prognosis in patients with diffuse large B-cell lymphoma: SUV-based assessment versus visual analysis. J Nucl Med. United States; 2007: 1626-32.
75. Shankar LK, Hoffman JM, Bacharach S, et al. Consensus recommendations for the use of 18F-FDG PET as an indicator of therapeutic response in patients in National Cancer Institute Trials. *J Nucl Med* 2006; **47**(6): 1059-66.
76. Kinahan PE, Fletcher JW. Positron emission tomography-computed tomography standardized uptake values in clinical practice and assessing response to therapy. *Seminars in ultrasound, CT, and MR* 2010; **31**(6): 496-505.
77. Vriens D, Visser EP, de Geus-Oei LF, Oyen WJ. Methodological considerations in quantification of oncological FDG PET studies. *European journal of nuclear medicine and molecular imaging* 2010; **37**(7): 1408-25.

78. Soret M, Bacharach SL, Buvat I. Partial-volume effect in PET tumor imaging. J Nucl Med. United States; 2007: 932-45.
79. Hoffman EJ, Huang SC, Phelps ME. Quantitation in positron emission computed tomography: 1. Effect of object size. *Journal of computer assisted tomography* 1979; **3**(3): 299-308.
80. Vesselle H, Schmidt RA, Pugsley JM, et al. Lung cancer proliferation correlates with [F-18]fluorodeoxyglucose uptake by positron emission tomography. *Clinical cancer research : an official journal of the American Association for Cancer Research* 2000; **6**(10): 3837-44.
81. Hallett WA, Marsden PK, Cronin BF, O'Doherty MJ. Effect of corrections for blood glucose and body size on [18F]FDG PET standardised uptake values in lung cancer. *Eur J Nucl Med* 2001; **28**(7): 919-22.
82. Weber WA. Use of PET for monitoring cancer therapy and for predicting outcome. J Nucl Med. United States; 2005: 983-95.
83. Wahl RL, Jacene H, Kasamon Y, Lodge MA. From RECIST to PERCIST: Evolving Considerations for PET response criteria in solid tumors. *J Nucl Med* 2009; **50 Suppl 1**: 122S-50S.
84. Boellaard R, Krak NC, Hoekstra OS, Lammertsma AA. Effects of noise, image resolution, and ROI definition on the accuracy of standard uptake values: a simulation study. J Nucl Med. United States; 2004: 1519-27.
85. Sher A, Lacoeuille F, Fosse P, et al. For avid glucose tumors, the SUV peak is the most reliable parameter for [(18)F]FDG-PET/CT quantification, regardless of acquisition time. *EJNMMI research* 2016; **6**(1): 21.
86. van den Hoff J. Principles of quantitative positron emission tomography. *Amino Acids* 2005; **29**(4): 341-53.
87. Hawkins RA, Choi Y, Huang SC, et al. Evaluation of the skeletal kinetics of fluorine-18-fluoride ion with PET. *J Nucl Med* 1992; **33**(5): 633-42.
88. Willemsen AT, van den Hoff J. Fundamentals of quantitative PET data analysis. *Current pharmaceutical design* 2002; **8**(16): 1513-26.
89. Siddique M, Frost ML, Blake GM, et al. The precision and sensitivity of (18)F-fluoride PET for measuring regional bone metabolism: a comparison of quantification methods. *J Nucl Med* 2011; **52**(11): 1748-55.
90. Blake GM, Park-Holohan SJ, Cook GJR, al. e. Quantitative studies of bone with the use of 18F-fluoride and 99mTc-methylene diphosphonate. *Semin Nucl Med* 2001; **31**: 28-49.
91. Cheng NM, Dean Fang YH, Tung-Chieh Chang J, et al. Textural Features of Pretreatment 18F-FDG PET/CT Images: Prognostic Significance in Patients with Advanced T-Stage Oropharyngeal Squamous Cell Carcinoma. *J Nucl Med* 2013; **54**(10): 1703-9.
92. Doot RK, Muzi M, Peterson LM, et al. Kinetic analysis of 18F-fluoride PET images of breast cancer bone metastases. *J Nucl Med* 2010; **51**(4): 521-7.
93. Greuter HN, Boellaard R, van Lingen A, Franssen EJ, Lammertsma AA. Measurement of 18F-FDG concentrations in blood samples: comparison of direct calibration and standard solution methods. *Journal of nuclear medicine technology* 2003; **31**(4): 206-9.
94. van der Weerdt AP, Klein LJ, Visser CA, Visser FC, Lammertsma AA. Use of arterialised venous instead of arterial blood for measurement of myocardial glucose metabolism during euglycaemic-hyperinsulinaemic clamping. *European journal of nuclear medicine and molecular imaging* 2002; **29**(5): 663-9.
95. Ohtake T, Kosaka N, Watanabe T, et al. Noninvasive method to obtain input function for measuring tissue glucose utilization of thoracic and abdominal organs. *J Nucl Med* 1991; **32**(7): 1432-8.
96. de Geus-Oei LF, Visser EP, Krabbe PF, et al. Comparison of image-derived and arterial input functions for estimating the rate of glucose metabolism in therapy-monitoring 18F-FDG PET studies. *J Nucl Med* 2006; **47**(6): 945-9.
97. Cook GJ, Lodge MA, Marsden PK, Dynes A, Fogelman I. Non-invasive assessment of skeletal kinetics using fluorine-18 fluoride positron emission tomography:

evaluation of image and population-derived arterial input functions. *Eur J Nucl Med* 1999; **26**(11): 1424-9.
98. Lee JR, Madsen MT, Bushnel D, Menda Y. A threshold method to improve standardized uptake value reproducibility. *Nucl Med Commun* 2000; **21**(7): 685-90.
99. Tomasi G, Turkheimer F, Aboagye E. Importance of quantification for the analysis of PET data in oncology: review of current methods and trends for the future. *Molecular imaging and biology : MIB : the official publication of the Academy of Molecular Imaging* 2012; **14**(2): 131-46.
100. Niyazi M, Landrock S, Elsner A, et al. Automated biological target volume delineation for radiotherapy treatment planning using FDG-PET/CT. *Radiation oncology* 2013; **8**(1): 180.
101. Nestle U, Kremp S, Schaefer-Schuler A, et al. Comparison of different methods for delineation of 18F-FDG PET-positive tissue for target volume definition in radiotherapy of patients with non-Small cell lung cancer. *J Nucl Med* 2005; **46**(8): 1342-8.
102. Foster B, Bagci U, Mansoor A, Xu Z, Mollura DJ. A review on segmentation of positron emission tomography images. *Comput Biol Med* 2014; **50c**: 76-96.
103. Ashamalla H, Rafla S, Parikh K, et al. The contribution of integrated PET/CT to the evolving definition of treatment volumes in radiation treatment planning in lung cancer. *International journal of radiation oncology, biology, physics* 2005; **63**(4): 1016-23.
104. Doll C, Duncker-Rohr V, Rucker G, et al. Influence of experience and qualification on PET-based target volume delineation. When there is no expert--ask your colleague. *Strahlentherapie und Onkologie : Organ der Deutschen Rontgengesellschaft [et al]* 2014; **190**(6): 555-62.
105. Buijsen J, van den Bogaard J, van der Weide H, et al. FDG-PET-CT reduces the interobserver variability in rectal tumor delineation. *Radiotherapy and oncology : journal of the European Society for Therapeutic Radiology and Oncology* 2012; **102**(3): 371-6.
106. Zaidi H, El Naqa I. PET-guided delineation of radiation therapy treatment volumes: a survey of image segmentation techniques. *European journal of nuclear medicine and molecular imaging* 2010; **37**(11): 2165-87.
107. Erdi YE, Mawlawi O, Larson SM, et al. Segmentation of lung lesion volume by adaptive positron emission tomography image thresholding. *Cancer* 1997; **80**(12 Suppl): 2505-9.
108. Miller TR, Grigsby PW. Measurement of tumor volume by PET to evaluate prognosis in patients with advanced cervical cancer treated by radiation therapy. *International journal of radiation oncology, biology, physics* 2002; **53**(2): 353-9.
109. Bradley J, Thorstad WL, Mutic S, et al. Impact of FDG-PET on radiation therapy volume delineation in non-small-cell lung cancer. *International journal of radiation oncology, biology, physics* 2004; **59**(1): 78-86.
110. Scarfone C, Lavely WC, Cmelak AJ, et al. Prospective feasibility trial of radiotherapy target definition for head and neck cancer using 3-dimensional PET and CT imaging. *J Nucl Med* 2004; **45**(4): 543-52.
111. Biehl KJ, Kong FM, Dehdashti F, et al. 18F-FDG PET definition of gross tumor volume for radiotherapy of non-small cell lung cancer: is a single standardized uptake value threshold approach appropriate? *J Nucl Med* 2006; **47**(11): 1808-12.
112. Ford EC, Kinahan PE, Hanlon L, et al. Tumor delineation using PET in head and neck cancers: threshold contouring and lesion volumes. *Medical physics* 2006; **33**(11): 4280-8.
113. Hatt M, Cheze le Rest C, Turzo A, Roux C, Visvikis D. A fuzzy locally adaptive Bayesian segmentation approach for volume determination in PET. *IEEE Trans Med Imaging* 2009; **28**(6): 881-93.
114. Hatt M, Cheze-Le Rest C, Aboagye EO, et al. Reproducibility of 18F-FDG and 3'-deoxy-3'-18F-fluorothymidine PET tumor volume measurements. *J Nucl Med* 2010; **51**(9): 1368-76.

115. Vallabhajosula S, Solnes L, Vallabhajosula B. A broad overview of positron emission tomography radiopharmaceuticals and clinical applications: what is new? *Semin Nucl Med* 2011; **41**(4): 246-64.
116. Beheshti M, Vali R, Waldenberger P, et al. Detection of bone metastases in patients with prostate cancer by 18F fluorocholine and 18F fluoride PET-CT: a comparative study. *European journal of nuclear medicine and molecular imaging* 2008; **35**(10): 1766-74.
117. Moertel CG, Hanley JA. The effect of measuring error on the results of therapeutic trials in advanced cancer. *Cancer* 1976; **38**(1): 388-94.
118. Miller AB, Hoogstraten B, Staquet M, Winkler A. Reporting results of cancer treatment. *Cancer* 1981; **47**(1): 207-14.
119. Thiesse P, Ollivier L, Di Stefano-Louineau D, et al. Response rate accuracy in oncology trials: reasons for interobserver variability. Groupe Francais d'Immunotherapie of the Federation Nationale des Centres de Lutte Contre le Cancer. *J Clin Oncol* 1997; **15**(12): 3507-14.
120. Hayward JL, Carbone PP, Heusen JC, Kumaoka S, Segaloff A, Rubens RD. Assessment of response to therapy in advanced breast cancer. *British journal of cancer* 1977; **35**(3): 292-8.
121. Therasse P, Arbuck SG, Eisenhauer EA, et al. New guidelines to evaluate the response to treatment in solid tumors. European Organization for Research and Treatment of Cancer, National Cancer Institute of the United States, National Cancer Institute of Canada. *Journal of the National Cancer Institute* 2000; **92**(3): 205-16.
122. Eisenhauer EA, Therasse P, Bogaerts J, et al. New response evaluation criteria in solid tumours: revised RECIST guideline (version 1.1). *Eur J Cancer* 2009; **45**(2): 228-47.
123. Young H, Baum R, Cremerius U, et al. Measurement of clinical and subclinical tumour response using [18F]-fluorodeoxyglucose and positron emission tomography: review and 1999 EORTC recommendations. European Organization for Research and Treatment of Cancer (EORTC) PET Study Group. *Eur J Cancer* 1999; **35**(13): 1773-82.
124. Boellaard R, Oyen WJ, Hoekstra CJ, et al. The Netherlands protocol for standardisation and quantification of FDG whole body PET studies in multi-centre trials. *European journal of nuclear medicine and molecular imaging* 2008; **35**(12): 2320-33.
125. Condon BR, Buchanan R, Garvie NW, et al. Assessment of progression of secondary bone lesions following cancer of the breast or prostate using serial radionuclide imaging. *Br J Radiol* 1981; **54**(637): 18-23.
126. Galasko CS. Diagnosis of skeletal metastases and assessment of response to treatment. *Clinical orthopaedics and related research* 1995; (312): 64-75.
127. Vogel CL, Schoenfelder J, Shemano I, Hayes DF, Gams RA. Worsening bone scan in the evaluation of antitumor response during hormonal therapy of breast cancer. *J Clin Oncol* 1995; **13**(5): 1123-8.
128. Cook GJ, Fogelman I. The role of nuclear medicine in monitoring treatment in skeletal malignancy. *Semin Nucl Med* 2001; **31**(3): 206-11.
129. Janicek MJ, Hayes DF, Kaplan WD. Healing flare in skeletal metastases from breast cancer. *Radiology* 1994; **192**(1): 201-4.
130. Coleman I. Bone Scan Flare Predicts Successful Systemic Therapy for Bone Metastases. *J Nucl Med* 1988; **29**: 1354-9.
131. Scher HI. Prostate carcinoma: defining therapeutic objectives and improving overall outcomes. *Cancer* 2003; **97**(3 Suppl): 758-71.
132. Scher HI, Halabi S, Tannock I, et al. Design and end points of clinical trials for patients with progressive prostate cancer and castrate levels of testosterone: recommendations of the Prostate Cancer Clinical Trials Working Group. *J Clin Oncol* 2008; **26**(7): 1148-59.
133. Vassiliou V, Kalogeropoulou C, Christopoulos C, Solomou E, Leotsinides M, Kardamakis D. Combination ibandronate and radiotherapy for the treatment of bone metastases: Clinical evaluation and radiologic assessment. *International Journal of Radiation Oncology*Biology*Physics* 2007; **67**(1): 264-72.

134. Wahl RL, Zasadny K, Helvie M, Hutchins GD, Weber B, Cody R. Metabolic monitoring of breast cancer chemohormonotherapy using positron emission tomography: initial evaluation. *J Clin Oncol* 1993; **11**(11): 2101-11.
135. Stafford SE, Gralow JR, Schubert EK, et al. Use of serial FDG PET to measure the response of bone-dominant breast cancer to therapy. *Academic radiology* 2002; **9**(8): 913-21.
136. Gwak HS, Youn SM, Chang U, et al. Usefulness of (18)F-fluorodeoxyglucose PET for radiosurgery planning and response monitoring in patients with recurrent spinal metastasis. *Minim Invasive Neurosurg* 2006; **49**(3): 127-34.
137. Du Y, Cullum I, Illidge TM, Ell PJ. Fusion of metabolic function and morphology: sequential [18F]fluorodeoxyglucose positron-emission tomography/computed tomography studies yield new insights into the natural history of bone metastases in breast cancer. *J Clin Oncol* 2007; **25**(23): 3440-7.
138. Tateishi U, Gamez C, Dawood S, Yeung HW, Cristofanilli M, Macapinlac HA. Bone metastases in patients with metastatic breast cancer: morphologic and metabolic monitoring of response to systemic therapy with integrated PET/CT. *Radiology* 2008; **247**(1): 189-96.
139. Lindholm P, Lapela M, Nagren K, Lehikoinen P, Minn H, Jyrkkio S. Preliminary study of carbon-11 methionine PET in the evaluation of early response to therapy in advanced breast cancer. *Nucl Med Commun* 2009; **30**(1): 30-6.
140. Cook G, Jr., Parker C, Chua S, Johnson B, Aksnes AK, Lewington VJ. 18F-fluoride PET: changes in uptake as a method to assess response in bone metastases from castrate-resistant prostate cancer patients treated with 223Ra-chloride (Alpharadin). *EJNMMI research* 2011; **1**(1): 4.
141. Chenevert TL, McKeever PE, Ross BD. Monitoring early response of experimental brain tumors to therapy using diffusion magnetic resonance imaging. *Clinical cancer research : an official journal of the American Association for Cancer Research* 1997; **3**(9): 1457-66.
142. Galons JP, Altbach MI, Paine-Murrieta GD, Taylor CW, Gillies RJ. Early increases in breast tumor xenograft water mobility in response to paclitaxel therapy detected by non-invasive diffusion magnetic resonance imaging. *Neoplasia* 1999; **1**(2): 113-7.
143. Theilmann RJ, Borders R, Trouard TP, et al. Changes in water mobility measured by diffusion MRI predict response of metastatic breast cancer to chemotherapy. *Neoplasia* 2004; **6**(6): 831-7.
144. Mardor Y, Roth Y, Ochershvilli A, et al. Pretreatment prediction of brain tumors' response to radiation therapy using high b-value diffusion-weighted MRI. *Neoplasia* 2004; **6**(2): 136-42.
145. Cui Y, Zhang XP, Sun YS, Tang L, Shen L. Apparent diffusion coefficient: potential imaging biomarker for prediction and early detection of response to chemotherapy in hepatic metastases. *Radiology* 2008; **248**(3): 894-900.
146. Thoeny HC, De Keyzer F, Chen F, et al. Diffusion-weighted MR imaging in monitoring the effect of a vascular targeting agent on rhabdomyosarcoma in rats. *Radiology* 2005; **234**(3): 756-64.
147. Manton DJ, Chaturvedi A, Hubbard A, et al. Neoadjuvant chemotherapy in breast cancer: early response prediction with quantitative MR imaging and spectroscopy. *British journal of cancer* 2006; **94**(3): 427-35.
148. Hayashida Y, Yakushiji T, Awai K, et al. Monitoring therapeutic responses of primary bone tumors by diffusion-weighted image: Initial results. *Eur Radiol* 2006; **16**(12): 2637-43.
149. Uhl M, Saueressig U, van Buiren M, et al. Osteosarcoma: preliminary results of in vivo assessment of tumor necrosis after chemotherapy with diffusion- and perfusion-weighted magnetic resonance imaging. *Investigative radiology* 2006; **41**(8): 618-23.
150. Lee KC, Sud S, Meyer CR, et al. An imaging biomarker of early treatment response in prostate cancer that has metastasized to the bone. *Cancer Res* 2007; **67**(8): 3524-8.

151. Lee KC, Bradley DA, Hussain M, et al. A feasibility study evaluating the functional diffusion map as a predictive imaging biomarker for detection of treatment response in a patient with metastatic prostate cancer to the bone. *Neoplasia* 2007; **9**(12): 1003-11.
152. Messiou C, Collins DJ, Giles S, de Bono JS, Bianchini D, de Souza NM. Assessing response in bone metastases in prostate cancer with diffusion weighted MRI. *Eur Radiol* 2011; **21**(10): 2169-77.
153. Reischauer C, Froehlich JM, Koh DM, et al. Bone metastases from prostate cancer: assessing treatment response by using diffusion-weighted imaging and functional diffusion maps--initial observations. *Radiology* 2010; **257**(2): 523-31.
154. Blackledge MD. Assessment of Treatment Response by Total Tumour Volume and Global Apparent Diffusion Coefficient Using Diffusion-Weighted MRI in Patients with Metastatic Bone Disease: A Feasibility Study. 2014.
155. Steinborn MM, Heuck AF, Tiling R, Bruegel M, Gauger L, Reiser MF. Whole-body bone marrow MRI in patients with metastatic disease to the skeletal system. *Journal of computer assisted tomography* 1999; **23**(1): 123-9.
156. Takahara T, Imai Y, Yamashita T, Yasuda S, Nasu S, Van Cauteren M. Diffusion weighted whole body imaging with background body signal suppression (DWIBS): technical improvement using free breathing, STIR and high resolution 3D display. *Radiation medicine* 2004; **22**(4): 275-82.
157. Ballon D, Watts R, Dyke JP, et al. Imaging therapeutic response in human bone marrow using rapid whole-body MRI. *Magnetic resonance in medicine : official journal of the Society of Magnetic Resonance in Medicine / Society of Magnetic Resonance in Medicine* 2004; **52**(6): 1234-8.
158. Nakanishi K, Kobayashi M, Nakaguchi K, et al. Whole-body MRI for detecting metastatic bone tumor: diagnostic value of diffusion-weighted images. *Magnetic resonance in medical sciences : MRMS : an official journal of Japan Society of Magnetic Resonance in Medicine* 2007; **6**(3): 147-55.
159. Xu X, Ma L, Zhang JS, et al. Feasibility of whole body diffusion weighted imaging in detecting bone metastasis on 3.0T MR scanner. *Chin Med Sci J* 2008; **23**(3): 151-7.
160. Takenaka D, Ohno Y, Matsumoto K, et al. Detection of bone metastases in non-small cell lung cancer patients: comparison of whole-body diffusion-weighted imaging (DWI), whole-body MR imaging without and with DWI, whole-body FDG-PET/CT, and bone scintigraphy. *Journal of magnetic resonance imaging : JMRI* 2009; **30**(2): 298-308.
161. Ghanem N, Lohrmann C, Engelhardt M, et al. Whole-body MRI in the detection of bone marrow infiltration in patients with plasma cell neoplasms in comparison to the radiological skeletal survey. *European Radiology* 2006; **16**(5): 1005-14.
162. Gutzeit A, Doert A, Froehlich JM, et al. Comparison of diffusion-weighted whole body MRI and skeletal scintigraphy for the detection of bone metastases in patients with prostate or breast carcinoma. *Skeletal Radiol* 2010; **39**(4): 333-43.
163. Wu LM, Gu HY, Zheng J, et al. Diagnostic value of whole-body magnetic resonance imaging for bone metastases: a systematic review and meta-analysis. *Journal of magnetic resonance imaging : JMRI* 2011; **34**(1): 128-35.
164. Eiber M, Holzapfel K, Ganter C, et al. Whole-body MRI including diffusion-weighted imaging (DWI) for patients with recurring prostate cancer: technical feasibility and assessment of lesion conspicuity in DWI. *Journal of magnetic resonance imaging : JMRI* 2011; **33**(5): 1160-70.
165. Haralick RM, Shanmugam K, Dinstein I. Textural Features for Image Classification. *IEEE Transactions on Systems, Man and Cybernetics* 1973; **3**(6): 610-21.
166. Soh L-K, Tsatsoulis C. Texture Analysis of SAR Sea Ice Imagery Using Gray Level Co-Occurrence Matrices. *IEEE Transactions on Geoscience and Remote Sensing* 1999; **37**(2): 780-95.
167. Narendra VG, Hareesh KS. Cashew Kernels Classification Using Texture Features. *International Journal of Machine Intelligence* 2011; **3**(2): 45-51.
168. Goldberger AL, West BJ. Fractals in physiology and medicine. *The Yale journal of biology and medicine* 1987; **60**(5): 421-35.

169. Daxer A. Characterisation of the neovascularisation process in diabetic retinopathy by means of fractal geometry: diagnostic implications. *Graefes Arch Clin Exp Ophthalmol* 1993; **231**: 681-6.
170. Claridge E, Hall PN, Keefe M, Allen JP. Shape analysis for classification of malignant melanoma. *Journal of biomedical engineering* 1992; **14**(3): 229-34.
171. Goh V, Ganeshan B, Nathan P, Juttla JK, Vinayan A, Miles KA. Assessment of Response to Tyrosine Kinase Inhibitors in Metastatic Renal Cell Cancer: CT Texture as a Predictive Biomarker. *Radiology* 2011; **261**(1): 165-71.
172. Michoux N, Van den Broeck S, Lacoste L, et al. Texture analysis on MR images helps predicting non-response to NAC in breast cancer. *BMC cancer* 2015; **15**(1): 574.
173. Miles KA, Ganeshan B, Griffiths MR, Young RC, Chatwin CR. Colorectal cancer: texture analysis of portal phase hepatic CT images as a potential marker of survival. *Radiology* 2009; **250**(2): 444-52.
174. Ganeshan B, Panayiotou E, Burnand K, Dizdarevic S, Miles K. Tumour heterogeneity in non-small cell lung carcinoma assessed by CT texture analysis: a potential marker of survival. *Eur Radiol* 2012; **22**(4): 796-802.
175. Ganeshan B, Skogen K, Pressney I, Coutroubis D, Miles K. Tumour heterogeneity in oesophageal cancer assessed by CT texture analysis: preliminary evidence of an association with tumour metabolism, stage, and survival. *Clinical radiology* 2012; **67**(2): 157-64.
176. Wibmer A, Hricak H, Gondo T, et al. Haralick texture analysis of prostate MRI: utility for differentiating non-cancerous prostate from prostate cancer and differentiating prostate cancers with different Gleason scores. *Eur Radiol* 2015; **25**(10): 2840-50.
177. Nketiah G, Elschot M, Kim E, et al. T2-weighted MRI-derived textural features reflect prostate cancer aggressiveness: preliminary results. *Eur Radiol* 2016.
178. Eary JF, O'Sullivan F, O'Sullivan J, Conrad EU. Spatial heterogeneity in sarcoma 18F-FDG uptake as a predictor of patient outcome. *J Nucl Med* 2008; **49**(12): 1973-9.
179. El Naqa I, Grigsby P, Apte A, et al. Exploring feature-based approaches in PET images for predicting cancer treatment outcomes. *Pattern recognition* 2009; **42**(6): 1162-71.
180. Tixier F, Le Rest CC, Hatt M, et al. Intratumor heterogeneity characterized by textural features on baseline 18F-FDG PET images predicts response to concomitant radiochemotherapy in esophageal cancer. *J Nucl Med* 2011; **52**(3): 369-78.
181. Nahmias C, Wahl LM. Reproducibility of standardized uptake value measurements determined by 18F-FDG PET in malignant tumors. *J Nucl Med* 2008; **49**(11): 1804-8.
182. Keyes JW, Jr. SUV: standard uptake or silly useless value? *J Nucl Med* 1995; **36**(10): 1836-9.
183. Contractor K, Challapalli A, Barwick T, et al. Use of [11C]choline PET-CT as a noninvasive method for detecting pelvic lymph node status from prostate cancer and relationship with choline kinase expression. *Clinical cancer research : an official journal of the American Association for Cancer Research* 2011; **17**(24): 7673-83.
184. Siddique M, Blake GM, Frost ML, et al. Estimation of regional bone metabolism from whole-body 18F-fluoride PET static images. *European journal of nuclear medicine and molecular imaging* 2012; **39**(2): 337-43.
185. Lim R, Eaton A, Lee NY, et al. 18F-FDG PET/CT metabolic tumor volume and total lesion glycolysis predict outcome in oropharyngeal squamous cell carcinoma. *J Nucl Med* 2012; **53**(10): 1506-13.
186. Hatt M, Tixier F, Cheze Le Rest C, Pradier O, Visvikis D. Robustness of intratumour (1)(8)F-FDG PET uptake heterogeneity quantification for therapy response prediction in oesophageal carcinoma. *European journal of nuclear medicine and molecular imaging* 2013; **40**(11): 1662-71.
187. Blackledge MD, Collins DJ, Tunariu N, et al. Assessment of treatment response by total tumor volume and global apparent diffusion coefficient using diffusion-weighted MRI in patients with metastatic bone disease: a feasibility study. *PLoS One* 2014; **9**(4): e91779.

188. Abouzied M, Demirkaya O, Rifai A. Quantification of bone metastasis in whole body images of 18 F-FDG PET/CT. *J NUCL MED MEETING ABSTRACTS* 2007; **48**(MeetingAbstracts_2): 144P-b-.
189. Tamada T, Kanomata N, Sone T, et al. High b value (2,000 s/mm2) diffusion-weighted magnetic resonance imaging in prostate cancer at 3 Tesla: comparison with 1,000 s/mm2 for tumor conspicuity and discrimination of aggressiveness. *PLoS One* 2014; **9**(5): e96619.
190. Hillner BE, Siegel BA, Hanna L, Duan F, Quinn B, Shields AF. 18F-fluoride PET used for treatment monitoring of systemic cancer therapy: results from the National Oncologic PET Registry. *J Nucl Med* 2015; **56**(2): 222-8.
191. Zukotynski KA, Kim CK, Gerbaudo VH, et al. (18)F-FDG-PET/CT and (18)F-NaF-PET/CT in men with castrate-resistant prostate cancer. *Am J Nucl Med Mol Imaging* 2015; **5**(1): 72-82.
192. Sanchez-Crespo A, Christiansson F, Thur CK, Lundblad H, Sundin A. Predictive value of [18F]-fluoride PET for monitoring bone remodeling in patients with orthopedic conditions treated with a Taylor spatial frame. *European journal of nuclear medicine and molecular imaging* 2017; **44**(3): 441-8.
193. Frost ML, Moore AE, Siddique M, et al. (1)(8)F-fluoride PET as a noninvasive imaging biomarker for determining treatment efficacy of bone active agents at the hip: a prospective, randomized, controlled clinical study. *Journal of bone and mineral research : the official journal of the American Society for Bone and Mineral Research* 2013; **28**(6): 1337-47.
194. Cheng C, Heiss C, Dimitrakopoulou-Strauss A, et al. Evaluation of bone remodeling with (18)F-fluoride and correlation with the glucose metabolism measured by (18)F-FDG in lumbar spine with time in an experimental nude rat model with osteoporosis using dynamic PET-CT. *Am J Nucl Med Mol Imaging* 2013; **3**(2): 118-28.
195. Apolo AB, Lindenberg L, Shih JH, et al. Prospective Study Evaluating Na18F PET/CT in Predicting Clinical Outcomes and Survival in Advanced Prostate Cancer. *J Nucl Med* 2016; **57**(6): 886-92.
196. Takalkar A, Paryani B, Adams S, Subbiah V. Radium-223 dichloride therapy in breast cancer with osseous metastases. *BMJ Case Rep* 2015; **2015**.
197. Sachpekidis C, Hillengass J, Goldschmidt H, et al. Treatment response evaluation with 18F-FDG PET/CT and 18F-NaF PET/CT in multiple myeloma patients undergoing high-dose chemotherapy and autologous stem cell transplantation. *European journal of nuclear medicine and molecular imaging* 2017; **44**(1): 50-62.
198. Yu EY, Duan F, Muzi M, et al. Castration-resistant prostate cancer bone metastasis response measured by 18F-fluoride PET after treatment with dasatinib and correlation with progression-free survival: results from American College of Radiology Imaging Network 6687. *J Nucl Med* 2015; **56**(3): 354-60.
199. Simoncic U, Perlman S, Liu G, Staab MJ, Straus JE, Jeraj R. Comparison of NaF and FDG PET/CT for assessment of treatment response in castration-resistant prostate cancers with osseous metastases. *Clinical genitourinary cancer* 2015; **13**(1): e7-e17.
200. Johns WD, Garnick MB, Kaplan WD. Leuprolide therapy for prostate cancer. An association with scintigraphic "flare" on bone scan. *Clinical nuclear medicine* 1990; **15**(7): 485-7.
201. Levenson RM, Sauerbrunn BJ, Bates HR, Newman RD, Eddy JL, Ihde DC. Comparative value of bone scintigraphy and radiography in monitoring tumor response in systemically treated prostatic carcinoma. *Radiology* 1983; **146**(2): 513-8.
202. Pollen JJ, Witztum KF, Ashburn WL. The flare phenomenon on radionuclide bone scan in metastatic prostate cancer. *AJR Am J Roentgenol* 1984; **142**(4): 773-6.
203. Cook GJR. PET and PET/CT imaging of skeletal metastases. *Cancer imaging : the official publication of the International Cancer Imaging Society* 2010; **10**(1): 1-8.
204. Etchebehere EC, Araujo JC, Fox PS, Swanston NM, Macapinlac HA, Rohren EM. Prognostic Factors in Patients Treated with 223Ra: The Role of Skeletal Tumor Burden on Baseline 18F-Fluoride PET/CT in Predicting Overall Survival. *J Nucl Med* 2015; **56**(8): 1177-84.

205. Lindgren Belal S, Sadik M, Kaboteh R, et al. 3D skeletal uptake of 18F sodium fluoride in PET/CT images is associated with overall survival in patients with prostate cancer. *EJNMMI research* 2017; **7**(1): 15.
206. Piccardo A, Puntoni M, Morbelli S, et al. 18F-FDG PET/CT is a prognostic biomarker in patients affected by bone metastases from breast cancer in comparison with 18F-NaF PET/CT. *Nuklearmedizin* 2015; **54**(4): 163-72.
207. Tomasi G, Shepherd T, Turkheimer F, Visvikis D, Aboagye E. Comparative assessment of segmentation algorithms for tumor delineation on a test-retest [(11)C]choline dataset. *Medical physics* 2012; **39**(12): 7571-9.
208. Daisne JF, Duprez T, Weynand B, et al. Tumor volume in pharyngolaryngeal squamous cell carcinoma: comparison at CT, MR imaging, and FDG PET and validation with surgical specimen. *Radiology* 2004; **233**(1): 93-100.
209. van Baardwijk A, Bosmans G, Boersma L, et al. PET-CT-based auto-contouring in non-small-cell lung cancer correlates with pathology and reduces interobserver variability in the delineation of the primary tumor and involved nodal volumes. *International journal of radiation oncology, biology, physics* 2007; **68**(3): 771-8.
210. Soloway MS, Hardeman SW, Hickey D, et al. Stratification of patients with metastatic prostate cancer based on extent of disease on initial bone scan. *Cancer* 1988; **61**(1): 195-202.
211. Wassberg C, Lubberink M, Sorensen J, Johansson S. Repeatability of quantitative parameters of 18F-fluoride PET/CT and biochemical tumour and specific bone remodelling markers in prostate cancer bone metastases. *EJNMMI research* 2017; **7**(1): 42.
212. Padhani AR, Gogbashian A. Bony metastases: assessing response to therapy with whole-body diffusion MRI. *Cancer imaging : the official publication of the International Cancer Imaging Society* 2011; **11 Spec No A**: S129-45.
213. Dzik-Jurasz A, Domenig C, George M, et al. Diffusion MRI for prediction of response of rectal cancer to chemoradiation. *Lancet* 2002; **360**(9329): 307-8.
214. Roth Y, Tichler T, Kostenich G, et al. High-b-value diffusion-weighted MR imaging for pretreatment prediction and early monitoring of tumor response to therapy in mice. *Radiology* 2004; **232**(3): 685-92.
215. Koh DM, Scurr E, Collins D, et al. Predicting response of colorectal hepatic metastasis: value of pretreatment apparent diffusion coefficients. *AJR Am J Roentgenol* 2007; **188**(4): 1001-8.
216. Sun YS, Zhang XP, Tang L, et al. Locally advanced rectal carcinoma treated with preoperative chemotherapy and radiation therapy: preliminary analysis of diffusion-weighted MR imaging for early detection of tumor histopathologic downstaging. *Radiology* 2010; **254**(1): 170-8.
217. King AD, Thoeny HC. Functional MRI for the prediction of treatment response in head and neck squamous cell carcinoma: potential and limitations. *Cancer imaging : the official publication of the International Cancer Imaging Society* 2016; **16**(1): 23.
218. Beets-Tan RGH, Beets GL. MRI for assessing and predicting response to neoadjuvant treatment in rectal cancer. *Nat Rev Gastroenterol Hepatol* 2014; **11**(8): 480-8.
219. Zhang G-Y, Wang Y-J, Liu J-P, et al. Pretreatment Diffusion-Weighted MRI Can Predict the Response to Neoadjuvant Chemotherapy in Patients with Nasopharyngeal Carcinoma. *BioMed research international* 2015; **2015**: 8.
220. Turkbey B, Shah VP, Pang Y, et al. Is apparent diffusion coefficient associated with clinical risk scores for prostate cancers that are visible on 3-T MR images? *Radiology* 2011; **258**(2): 488-95.
221. Humphrey PA. Gleason grading and prognostic factors in carcinoma of the prostate. *Modern pathology : an official journal of the United States and Canadian Academy of Pathology, Inc* 2004; **17**(3): 292-306.
222. Sabbatini P, Larson SM, Kremer A, et al. Prognostic significance of extent of disease in bone in patients with androgen-independent prostate cancer. *J Clin Oncol* 1999; **17**(3): 948-57.

223. Dennis ER, Jia X, Mezheritskiy IS, et al. Bone scan index: a quantitative treatment response biomarker for castration-resistant metastatic prostate cancer. J Clin Oncol. United States; 2012: 519-24.
224. Fehr D, Veeraraghavan H, Wibmer A, et al. Automatic classification of prostate cancer Gleason scores from multiparametric magnetic resonance images. *Proceedings of the National Academy of Sciences of the United States of America* 2015; **112**(46): E6265-73.
225. Nilsen L, Fangberget A, Geier O, Olsen DR, Seierstad T. Diffusion-weighted magnetic resonance imaging for pretreatment prediction and monitoring of treatment response of patients with locally advanced breast cancer undergoing neoadjuvant chemotherapy. *Acta Oncologica* 2010; **49**(3): 354-60.
226. Leboulleux S, Dromain C, Vataire AL, et al. Prediction and diagnosis of bone metastases in well-differentiated gastro-entero-pancreatic endocrine cancer: a prospective comparison of whole body magnetic resonance imaging and somatostatin receptor scintigraphy. *J Clin Endocrinol Metab* 2008; **93**(8): 3021-8.
227. Chenevert TL, Stegman LD, Taylor JM, et al. Diffusion magnetic resonance imaging: an early surrogate marker of therapeutic efficacy in brain tumors. *Journal of the National Cancer Institute* 2000; **92**(24): 2029-36.
228. Fleten KG, Bakke KM, Abildgaard A, Mælandsmo GM, Redalen KR, Flatmark K. Abstract 4218: Diffusion-weighted MRI can predict response to aflibercept in in vivo models. *Cancer Research* 2016; **76**(14 Supplement): 4218-.
229. Horger M, Weisel K, Horger W, Mroue A, Fenchel M, Lichy M. Whole-body diffusion-weighted MRI with apparent diffusion coefficient mapping for early response monitoring in multiple myeloma: preliminary results. *AJR Am J Roentgenol* 2011; **196**(6): W790-5.
230. Saip P, Tenekeci N, Aydiner A, et al. Response evaluation of bone metastases in breast cancer: value of magnetic resonance imaging. *Cancer investigation* 1999; **17**(8): 575-80.
231. Sharma U, Danishad KKA, Seenu V, Jagannathan NR. Longitudinal study of the assessment by MRI and diffusion-weighted imaging of tumor response in patients with locally advanced breast cancer undergoing neoadjuvant chemotherapy. *NMR in biomedicine* 2009; **22**(1): 104-13.
232. Graham TJ, Box G, Tunariu N, et al. Preclinical evaluation of imaging biomarkers for prostate cancer bone metastasis and response to cabozantinib. *Journal of the National Cancer Institute* 2014; **106**(4): dju033.
233. Rozel S, Galban CJ, Nicolay K, et al. Synergy between anti-CCL2 and docetaxel as determined by DW-MRI in a metastatic bone cancer model. *Journal of cellular biochemistry* 2009; **107**(1): 58-64.
234. Messiou C, deSouza NM. Diffusion Weighted Magnetic Resonance Imaging of metastatic bone disease: A biomarker for treatment response monitoring. *Cancer biomarkers : section A of Disease markers* 2010; **6**(1): 21-32.
235. Cook GJ, Houston S, Rubens R, Maisey MN, Fogelman I. Detection of bone metastases in breast cancer by 18FDG PET: differing metabolic activity in osteoblastic and osteolytic lesions. *J Clin Oncol* 1998; **16**(10): 3375-9.
236. Nakajo M, Kajiya Y, Kaneko T, et al. FDG PET/CT and diffusion-weighted imaging for breast cancer: prognostic value of maximum standardized uptake values and apparent diffusion coefficient values of the primary lesion. *European journal of nuclear medicine and molecular imaging* 2010; **37**(11): 2011-20.
237. Yun TJ, Kim JH, Kim KH, Sohn CH, Park SW. Head and neck squamous cell carcinoma: differentiation of histologic grade with standard- and high-b-value diffusion-weighted MRI. *Head Neck* 2013; **35**.
238. Wang J, Takashima S, Takayama F, Kawakami S, Saito A, Matsushita T. Head and neck lesions: characterization with diffusion-weighted echo-planar MR imaging. *Radiology* 2001; **220**.
239. Ahn SJ, Choi SH, Kim YJ, Kim KG, Sohn CH, Han MH. Histogram analysis of apparent diffusion coefficient map of standard and high B-value diffusion MR imaging in

head and neck squamous cell carcinoma: a correlation study with histological grade. *Academic radiology* 2012; **19**.
240. Perez-Lopez R, Lorente D, Blackledge MD, et al. Volume of Bone Metastasis Assessed with Whole-Body Diffusion-weighted Imaging Is Associated with Overall Survival in Metastatic Castration-resistant Prostate Cancer. *Radiology* 2016; **280**(1): 151-60.
241. Vargas HA, Wassberg C, Fox JJ, et al. Bone Metastases in Castration-Resistant Prostate Cancer: Associations between Morphologic CT Patterns, Glycolytic Activity, and Androgen Receptor Expression on PET and Overall Survival. *Radiology* 2014: 130625.
242. Kaboteh R, Damber JE, Gjertsson P, et al. Bone Scan Index: a prognostic imaging biomarker for high-risk prostate cancer patients receiving primary hormonal therapy. *EJNMMI research* 2013; **3**(1): 9.
243. Altman DG. MEDICINE AND MATHEMATICS - STATISTICS AND ETHICS IN MEDICAL-RESEARCH .3. HOW LARGE A SAMPLE. *British Medical Journal* 1980; **281**(6251): 1336-8.
244. Altman DG. Practical statistics for medical research. Boca Raton, Fla.: Chapman & Hall/CRC; 1999.

Chapter 17 Appendix 1 - Trial Design

17.1 Introduction

The results reported in this thesis are the preliminary analyses of data from a family of three prospective complementary functional imaging studies (Functional Assessment of Bone metastases (FAB) – Prostate (FAB-P), Breast (FAB-B), and Integrin Expression (FAB-IE)). These are, at the time of writing, still recruiting patients. These three studies are aiming to identify functional imaging biomarkers of bone metastases to predict treatment response earlier than is currently achievable with standard clinical and radiological approaches.

For completeness, the following chapter summaries the three studies from which data for this thesis has drawn. The aims and objectives of these studies is not the primary focus of this thesis and will be achieved after full recruitment has been met.

17.2 FAB-P Methods

17.2.1 FAB-P Overview

FAB-P (Functional Assessment of Bone metastases – Prostate) is a prospective exploratory non-randomised study, aiming to recruit 34 patients with confirmed progressive bone predominant metastatic prostate cancer. At baseline, all patients will receive a whole-body MRI with DW-MRI, a dynamic localised ^{11}C-choline PET/CT scan, a static whole-body ^{11}C-Choline PET/CT scan and a static whole-body ^{18}F-Fluoride PET/CT scan before starting a new line in systemic chemotherapy with docetaxel. Patients will complete a clinical pain assessment questionnaire and have serum measurement of PSA, ALP and bone turn over markers. Patients will have had imaging of their bone disease (either a bone scan or CT) to determine PD as part of their routine care. The imaging and clinical assessments will be repeated eight weeks after starting the new treatment, with further pain assessment and blood tests at 12 and 24 weeks after initiation of therapy. If patients have increased activity on the week-8 ^{18}F-Fluoride scan suggesting a possible flare response, a further ^{18}F-Fluoride PET/CT scan at 12 weeks will be indicated for further assessment. FAB-P has Research Ethics Committee approval (NRES Committee South East Coast, REC reference 12/LO/0830) and ARSAC approval (certificate reference number 261/3186/28833)

17.2.2 FAB-P Objectives

The primary objective of FAB-P is to identify a treatment response or PD earlier (at 8 weeks) than current clinical and radiological assessment (PCWG2 guidelines) using DW-MRI and PET biomarkers. Secondarily FAB-P will be used to determine the diagnostic accuracy of each functional imaging method in staging the skeleton before treatment starts, and enable exploration of difference in baseline measurements from each method between subsequent responders and non-responders. Correlation of changes between scans with changes in serum biomarkers (PSA, ALP and bone turn over markers) will be sought.

17.2.3 FAB-P Planned recruitment

Sample size estimation is important to determine how many participants are necessary to adequately interrogate the study hypothesis. If too few participants are recruited the results may not achieve statistical significance; if too many, this is an unnecessary use of trial resources and study participants. Several factors contribute to the required sample size. For a study measuring therapeutic response it is necessary to consider the size of response that might be clinically relevant, often referred to as ∂. The smaller the change to be identified, the larger the required sample size necessary to prove the change is real, and not simply random error. There needs to be a practical balance struck; too large a difference required to identify a 'response' with the technique or treatment under investigation may limit the clinical usefulness of the study, and make it an unattainably large target to achieve.

Alman[243] suggested a nomogram to help select an appropriate sample size. He suggests the use of a 'standardised difference', a standardised measure of the relevant difference that might be identified. For example, in a PET response study it might be deemed clinically relevant to expect a reduction in SUV of 3 Bq/ml in participants who have a treatment response. Previous experience might indicate that the standard deviation (s) of SUV uptake is 2 Bq/ml, and thus the standardised difference can be calculated by 2/3 (s/∂). A decision is needed as to what significance level is required, i.e. the probability that a finding is true (often referred to as α). This is usually expected to be a two-sided significance level, representing the notion that the result might change in one of two directions, either getting larger or smaller. The power of a study is the probability that the outcome is not a false negative. If the probability of a false negative is β, power=1-β. Accepted academic convention normally requires a significance level of 0.05 or lower, but if multiple variables are to be measured it might be necessary to plan for a lower significance level. A power of between 80-90% is normally required[243]. The greater the required significance level and the power, the larger sample size is necessary.

Particularly in early exploratory research the standard deviation of the variables may not be known before the study starts. Altman suggests using the initial data from a study to identify the standard deviation before finalising the required sample size[244]. This approach, however, is not suitable for modern clinical studies requiring extensive regulatory oversight and pre-planning. He also

suggests that a standardised difference of 1.0 can be used, signifying a difference of greater than 1 standard deviation, irrespective of the actual value of the standard deviation[244].

At the time of writing clinically relevant differences in PET or MRI quantification parameters are not clearly defined in the literature that might correspond with an assessment of treatment response.

Therefore, aiming for a power of 80%, a statistical significance of p=0.05, and an assumed standardised difference of 1.0, it is possible to read from Altman's nomogram that a study group of around 30 is required. It was therefore concluded that a study group of 34 was required, to allow some leeway for participants who may withdraw from the study.

This technique could be criticised because of the number of variables to be examined and the different imaging techniques to be interrogated within the same study. A Bonferroni correction for multiple variables would result in a lower p-value for statistical significance. However, this would significantly increase the number of patients required. This research is preliminary research into the response assessment of bone metastases, and therefore a practical study group size has been necessary.

17.2.4 FAB-P Inclusion Criteria

Only men over 18 years of age with histologically confirmed prostate cancer, and with active skeletal-predominant disease, who are treatment naïve or have PD, and are to embark on new docetaxel chemotherapy, who are able and willing to give written informed consent, and who are willing and able to comply with the scheduled visits and tests, will be included in the study.

17.2.5 FAB-P Exclusion Criteria

Patients with uncontrolled concomitant medical conditions, contraindications to MR imaging, or likely to require palliative radiotherapy within the first 8 weeks of treatment, likely to require G-CSF support, or those commencing on concomitant bisphosphonate therapy, will be excluded from the study. Bisphosphonates have anti-osteoclastic, and possibly anti-tumour, function, and were therefore excluded to limit response assessment to anti-tumour therapies.

17.2.6 FAB-P PET Scan Procedure

Both PET scans can be completed on the same day due to the short half-life of ^{11}C. Patients will be fasted for 4 hours prior to the ^{11}C-choline PET/CT. 740MBq of ^{11}C-choline will be administered, and a dynamic PET scan will be acquired over a target bone lesion identified from recent imaging (if possible the target lesion will be selected to allow inclusion of the mediastinal blood pool in the field of view.) A whole-body ^{11}C-choline PET scan will then be obtained from mid-thigh to skull base, starting 10 minutes after injection, and after a low-dose CT scan for attenuation correction and image fusion. Venous blood samples (to measure plasma activity) will be taken at 11 and 40 (exact timings will be recorded; see SOP in Appendix).

60 minutes after the ^{11}C-Choline injection, 250MBq of ^{18}F-Fluoride tracer will be administered. ^{18}F-fluoride PET scans will be acquired from mid-thigh to the skull base at 60 minutes after tracer injection. Venous blood samples will be taken at 55 and 90 minutes after ^{18}F-Fluoride injection (i.e. before and after the whole-body scan) (please see SOP in Appendix).

PET scanning will be repeated 8 weeks after initiation of the new systemic therapy. The same bone lesion as baseline imaging will be imaged for the dynamic component of the ^{11}C-Choline PET. If patients are identified as having a possible flare-response on the 8 week ^{18}F-Fluoride scan a further 18F-Fluoride PET/CT can be carried out at 12 weeks.

17.2.7 FAB-P MRI Procedure

No specific patient preparation is necessary. Whole-Body MRI with DW-MRI will be performed using a 1.5T MR imaging system (Echo-planar spin-echo technique, free breathing, STIR fat suppression, with two b-values of 0 and 800 s/mm^2, and patients will be scanned from skull base to mid-thigh). Repeat DW-MRI imaging will be undertaken 8 weeks after initiation of the systemic therapy.

17.2.8 FAB-P Response Assessment

17.2.9 FAB-P Clinical Response Assessment

The patients will be recruited from the treatment centre prostate cancer clinics; all therapies and clinical follow up assessments and investigations are at the discretion of the clinical teams.

A response assessment will be made up to 24 weeks utilising the Prostate Cancer Working Group (PCWG2) criteria for defining PD, alongside the clinical judgement of the clinical teams primarily involved with the patient. The clinical teams, blinded to the results of the trial scans, will assess whether the patient has clinically responded or not. PD by 24 weeks will be guided by the PCWG2 criteria (see table below).

Feature	Definition of Progression
PSA	*If PSA decline from baseline* • ≥25% and ≥2ng/mL above nadir, confirmed at least 3 weeks later *If no PSA decline from baseline* • ≥25% and ≥2ng/mL after 12 weeks (NB - Ignore early rises in PSA before 12 weeks of treatment, unless other evidence of progression)
Soft Tissue Lesions	As RECIST criteria
Bone Lesions	New lesions confirmed on repeat scan at least 6 weeks later
Symptoms	Clinical suspicion confirmed by continuation/progression of symptoms with second review at least 3 weeks later

Table 17-1: Summary of PCGW2 Guidelines for defining PD[132]

A response to treatment will be defined as non-progression. There is no limitation within this study as to when response assessments or therapy changes might be made by the clinical teams

17.2.10 FAB-P PET - Response Assessment

At baseline, an independent reviewer (an experienced nuclear medicine physician), blinded to the conventional staging results will assess each scan, and record the disease stage and all sites of disease involvement, documenting the number and location of visible sites of disease that show increased metabolic activity greater than the background normal tissue levels. The largest and/or hottest 5 lesions will undergo further analysis with measurement of SUV_{mean} and SUV_{max} for each. The mean values from the 5 lesions will be used as the overall semi-quantitative score of metabolic activity in each patient, and will be compared between baseline and 8 weeks.

PERCIST response criteria will be used to classify treatment response on the PET images[83]:

1. Partial Metabolic Response (PMR) = >30% reduction in SUL (SUV corrected to lean body mass) peak
2. Stable Metabolic Disease (SMD) = Neither complete metabolic response, PMD nor SMD
3. Progressive Metabolic Disease (PMD) = >30% increase in SUL peak (with at least absolute rise of 0.8SUL) OR visible increase in uptake OR new metabolically active lesions

The parameter Ki will also be measured for each representative lesion using the ^{18}F-fluoride PET data.

17.2.11 FAB-P DW-MRI - Response Assessment

At baseline, independent reviewers blinded to the conventional staging results will assess each scan, and record sites of metastatic disease using a confidence score and subsequently corroborated with all available imaging findings. An ADC map will be created for each scan, and for response analysis differences in mean ADC values between pre-treatment and post-treatment will be recorded. ADC values will be calculated on a per patient and a per lesion basis to include up to the largest 5 lesions per patient. The ADC_{mean} value will be recorded for each of the five lesions, and the average of these will be recorded. For each patient and lesion the absolute and percentage difference in ADC_{mean} between baseline and post-treatment will be calculated.

Progressors will be defined by a greater than 12% decrease in mean ADC values. Stable disease will be represented by ADC values between baseline and a 12% decrease. Receiver operator characteristic analysis will be used to identify a threshold value that best dichotomises patients into response groups.

17.2.12 FAB-P End Points

- Clinical/radiological response (as per PCWG2 guidelines) at up to 24 weeks
- ^{11}C-Choline response (SUV_{max}, SUV_{mean}, SUL_{peak}) at 8 weeks
- ^{18}F-Fluoride response (SUV_{max}, SUV_{mean}, SUL_{peak}, Ki) at 8 weeks
- DW-MRI response (ADC_{mean}) at 8 weeks

17.3 FAB-IE Methods

17.3.1 FAB-IE Overview

FAB-IE (Functional Assessment of Bone metastases – Integrin Expression) is a prospective exploratory non-randomised study aiming to recruit 75 patients with confirmed progressive bony metastatic disease from prostate cancer (25 patients), breast cancer (25 patients), and progressive multiple myeloma (25 patients). Patients who meet the inclusion criteria will be considered for recruitment if they are to be started on a new course of conventional systemic therapy. For prostate cancer, this treatment will be Abiraterone. For breast cancer and multiple myeloma patients this will be a chemotherapy regimen at the discretion of the treating oncologist. At baseline, before treatment starts, all patients will receive a whole-body DW-MRI scan, a DCE-MRI of a specified target lesion, and a ^{99m}Tc-Maraciclatide SPECT-CT (consisting of a dynamic scan over the specified target lesion, a whole-body planar scan and a localised tomographic SPECT-CT of the target lesion). All patients will also have a clinical pain assessment questionnaire to complete, and will have blood tests taken for serum ALP, CTX, TRAcP-5b, cathepsin-K and circulating tumour cell measurement. These assessments will all be repeated 12 weeks after the patients starts treatment.

17.3.2 FAB-IE Planned recruitment

For the whole FAB-IE study, requiring a power of 80%, a statistical significance level of p=0.05 (two sided), a study population of 75 participants would equate to a clinically relevant standardised difference of 0.65 i.e. standard deviation of variable/clinically relevant difference[243]. The study design stipulated, therefore, 25 patients from each tumour group (prostate cancer, breast cancer, and myeloma). For the prostate cancer patients alone, 25 patients will equate to standardised difference of just over 1 to equate with a power of 80% and statistical significance of p=0.05, meaning that a change of variable greater than 1 standard deviation from the mean will be used to define a clinically significant change.

17.3.3 FAB-IE Inclusion Criteria

Only patients (over 18 years of age) with histologically confirmed prostate cancer, breast cancer or multiple myeloma, and with active skeletal metastatic disease (with at least 1 lesion >2cm), and who are to embark on new systemic therapy and are able and willing to give written informed consent and comply with the scheduled visits, will be included in the study.

17.3.4 FAB-IE Exclusion Criteria

Pregnant or lactating women will be excluded from the study. Also, excluded will be patients with uncontrolled concomitant medical conditions, those with contraindications to MR Imaging, those likely to require palliative radiotherapy within the first 12 weeks of treatment, those likely to require G-CSF support during treatment, and also those with a prognosis of<6 months.

17.3.5 FAB-IE SPECT/CT Scan Procedure

There will be three components to the SPECT/CT scan:

1. A dynamic planar acquisition (10-second frames, for 300 seconds following injection of the tracer 99mTc-maraciclatide) of a target metastatic site derived from the pre-treatment imaging (the same target lesion to be also used for MRI imaging)
2. Whole-body planar scan (45 minutes after injection)
3. SPECT-CT acquisition of target lesion (75 minutes after injection)

17.3.6 FAB-IE DCE-MRI Procedure

No specific patient preparation is necessary. DCE-MRI will be performed of the same target metastatic site identified for SPECT-CT. This will be performed using a 1.5T imaging system.

17.3.7 FAB-IE DW-MRI Procedure

No specific patient preparation is necessary. DW-MRI will be performed using a 1.5T MR imaging system (Echo-planar spin-echo technique, free breathing, STIR fat suppression, with two b-values of 0 and 800 s/mm^2, and patients will be scanned from skull base to mid-thigh.

17.3.8 FAB-IE Response Assessment

FAB-IE Clinical Response Assessment

Patients will be assessed by 2 specified oncologists/haematologists, blinded to the study scan results, at 12 and 24 weeks with radiological assessment using standard radiological approaches as felt clinically relevant, unless such investigations are clinically indicated before these times. Response will be based on clinical/biochemical measures

For prostate cancer patients, PD will be concluded if there is worsening of symptoms, a need for palliative radiotherapy, a rise in tumour markers or ALP, or worsening imaging per the PCWG2 guidance.

RECIST imaging response criteria will be used to define treatment response in breast cancer using metastases in soft tissue rather than bone. Progression may also be concluded if there is worsening of symptoms, a need for palliative radiotherapy, a rise in ALP or confirmed new sites of bony metastatic disease.

For those patients recruited on the multiple myeloma arm, a standard assessment of response will be completed as clinically indicated (para-protein or free light chain measurement, haemoglobin level, or bone marrow assessment).

FAB-IE SPECT - Response Assessment

The perfusion index of the target lesion will be derived from the lesion uptake curve compared to the contralateral normal region of interest (Hermes Gold nuclear medicine analysis software, Sweden). The number and distribution of skeletal metastases will be documented.

Uptake quantification of up to 5 marker bone metastases larger than 2cm will be included, measuring the target-to-normal ROI ratio and the target-to-liver ratio. Parameters from the whole-body planar images and the SPECT-CT acquisition will be acquired.

These parameters will be compared between baseline and the week 12 scan.

FAB-IE DCE-MRI - Response Assessment

AUC and K^{trans} will be calculated at baseline and at 12 weeks.

FAB-IE DW-MRI - Response Assessment

ADC values will be compared between baseline and the 12-week scan.

FAB-IE Serum Markers - Response Assessment

TRAcP-5b, cathepsin-K, CTX and ALP blood levels at baseline and at 12 weeks will be compared to measure osteoblastic activity and global skeletal resorption.

PSA changes will be recorded at baseline and at 12 weeks for prostate cancer patients. Myeloma patients will have paraprotein levels/free light chain levels, haemoglobin and bone marrow data correlated with imaging outcomes.

Circulating tumour cell numbers will be recorded at baseline and at 12 weeks.

17.3.9 FAB-IE End Points

Clinical/radiological response up to 24 weeks.

Level and variability of alphavbeta3 integrin expression in bone metastases in patients with breast cancer, prostate cancer or multiple myeloma.

Changes in alphavbeta3 integrin expression (measured at 12 weeks) in bone metastases in responders and non-responders (defined at up to 24 weeks)

Changes in DCE/DW MRI parameters of vascularity and cellularity

Changes in osteoclastic activity as measured by serum TRAcP-5b, cathepsin-K and bone resorption measured by CTX (measured at 12 weeks, compared with clinical/radiological response at up to 24 weeks)

Changes in circulating tumour cell measurements (measured at 12 weeks, compared with clinic-radiological response at up to 24 weeks)

17.4 FAB-B Methods

17.4.1 FAB-B Overview

FAB-B (Functional Assessment of Bone metastases – Breast) is a prospective exploratory non-randomised study, aiming to recruit 34 patients with confirmed progressive bony metastatic breast cancer. The patients will undergo whole-body 18F-Fluoride PET/CT, a whole-body 18F-FDG PET/CT and a whole-body DW-MRI before starting a new line in systemic chemotherapy. Patients will also complete a clinical pain assessment and quality of life measure, and will have serum measurement of ALP and bone turn over markers. The imaging and clinical assessments will be repeated eight weeks after starting the new treatment, with further pain assessment, quality of life and blood tests at 12 and 24 weeks after initiation of systemic therapy. If patients have a suspected flare response on the 8 week 18F-FDG PET/CT scan a further scan will be repeated at 12 weeks from baseline.

17.4.2 FAB-B Objectives

The primary objective of FAB-B is to identify a treatment response or PD earlier (at 8 weeks) than current clinical and radiological assessment at up to 24 weeks using DW-MRI and PET biomarkers. Secondary objectives include determining the diagnostic accuracy of the functional imaging methods in staging the skeleton before treatment starts, and to enable exploration of the difference in baseline functional biomarkers between those who subsequently respond and progress on therapy. It is also planned to explore differences in the functional imaging biomarkers between different metastatic phenotypes of bone metastases, i.e. sclerotic compared with lytic disease, and how the imaging biomarkers change with therapy, perhaps indicating the likely mechanisms involved with MRI signal changes.

17.4.3 FAB-B Planned Recruitment

Using the same methodology as for FAB-P (see earlier), FAB-B has been designed to require 30 patients to achieve a power of 80% with a significance level of 0.05. The planned recruitment is 34 to allow for some attrition during the study.

17.4.4 FAB-B Inclusion Criteria

Only women over 18 years will be included in the study, and they will be required to have confirmed skeletal predominant metastatic breast cancer who are either treatment naïve or embarking on new systemic chemotherapy or endocrine therapy alongside a bone targeted therapy (denosumab or a bisphosphonate). Participants will need to have a willingness to comply with the necessary study visits and tests.

17.4.5 FAB-B Exclusion Criteria

Pregnant or lactating women will be excluded from recruitment, as will those patients with concomitant uncontrolled medical conditions, poorly controlled diabetes mellitus, contraindications to PET or MRI imaging, and those patients who are likely to require palliative radiotherapy to bone metastases within the first 8 weeks of therapy. Patients likely to require GCS-F were also excluded.

17.4.6 FAB-B PET Scan Procedure

The half-life of ^{18}F requires the FDG and Fluoride PET scans to be completed on separate days to ensure no image contamination; a minimum of 12 hours has been stipulated for this study.

Patients will fast for 4 hours prior to the ^{18}F-FDG PET/CT scan. 400MBq of ^{18}F-FDG will be administered and a whole-body scan from mid-thigh to skull base will be started 60 minutes after administration of the tracer, and take 30 minutes to acquire.

On a separate day patients, will be prepared for the ^{18}F-Fluoride PET/CT scan in a similar way, with a 4 hour fast. 150MBq of 18F-Fluoride tracer will be administered, and a whole-body scan (mid-thigh to skull-base) will be acquired at 60 minutes after tracer administration, taking 30 minutes to scan. Venous blood samples will be taken at 30 minutes and 60 minutes after tracer injection (i.e. before and after the whole-body scan) to enable calculation of Ki.

17.4.7 FAB-B DW-MRI Scan Procedure

No specific patient preparation is necessary. Whole-Body DW-MRI will be performed using a 1.5T MR imaging system (Echo-planar spin-echo technique, free breathing, STIR fat suppression, with two b-values of 0 and 800 s/mm^2, and patients will be scanned from skull base to mid-thigh). Repeat DW-MRI imaging will be undertaken 8 weeks after initiation of the systemic therapy.

17.4.8 FAB-B Response Assessment

FAB-B Clinical Response Assessment

The patients will be recruited from the treatment centre prostate cancer clinics; all therapies and clinical follow up assessments and investigations are at the discretion of the clinical teams.

Assessment will be made by the patient's oncologist, using clinical assessment and other imaging methods guided by the oncologist's normal practice. The treating oncology team will be blinded to the results of the trial imaging. A clinical response assessment will be made by 24 weeks.

FAB-B PET – Response Assessment

At baseline, independent reviewers blinded to the conventional staging results will assess each scan, and record the disease stage and all sites of disease involvement, documenting the number and location of visible sites of disease that show increased metabolic activity greater than the background normal tissue levels. The largest 5 lesions will undergo further analysis with measurement of SUV_{mean} and SUV_{max} for each. The mean values from the 5 lesions will be used as the overall semi-quantitative score of metabolic activity in each patient, and will be compared between baseline and 8 weeks.

PERCIST response criteria will be used to classify treatment response on the PET images <Wahl, 2009 #932>:

1. Partial Metabolic Response (PMR) = >30% reduction in SUL (SUV corrected to lean body mass) peak
2. Stable Metabolic Disease (SMD) = Neither complete metabolic response, PMD nor SMD
3. Progressive Metabolic Disease (PMD) = >30% increase in SUL peak (with at least absolute rise of 0.8SUL) OR visible increase in uptake OR new metabolically active lesions

The parameter Ki will also be measured for each representative lesion using the ^{18}F-fluoride PET data.

FAB-B DW-MRI – Response Assessment

At baseline, independent reviewers blinded to the conventional staging results will assess each scan, and record sites of metastatic disease using a confidence score and subsequently corroborated with all available imaging findings. An ADC map will be created for each scan, and for response analysis differences in mean ADC values between pre-treatment and post-treatment will be recorded. ADC values will be calculated on a per patient and a per lesion basis to include up to the largest 5 lesions per patient. The mean ADC value will be recorded for each of the five lesions, and the average of these will be recorded. For each patient and lesion the absolute and percentage difference in mean ADC between baseline and post-treatment will be calculated. Progressors will be defined by a greater than 12% increase in mean ADC values. Stable disease will be represented by ADC values between baseline and a 12% increase. Receiver operator characteristic analysis will be used to identify a threshold value that best dichotomises patients into response groups.

17.4.9 FAB-B End Points

- Clinical/radiological response at up to 24 weeks
- ^{18}F-FDG response (SUV_{max}, SUV_{mean}) at 8 weeks
- ^{18}F-Fluoride response (SUV_{max}, SUV_{mean}, Ki) at 8 weeks

DW-MRI response (ADC) at 8 weeks